Machines That Learn:

Aesthetics of Adaptive Behaviors in Agent-based Art

Jean-Sébastien Senécal

A thesis

in

The Department

of

Interdisciplinary Studies

Presented in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

Concordia University

Montréal, Québec, Canada

September 2016

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Mr. Jean-Sébastien Senécal**

Entitled: **Machines That Learn: Aesthetics of Adaptive Behaviors in Agent-based Art**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Ph. D. in Humanities)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining commitee:

______________________________ Chair

______________________________ External Examiner

______________________________ Examiner

______________________________ Examiner

______________________________ Examiner

______________________________ Supervisor

Approved by: ______________________________

Dr Bina Freiwald, Graduate Program Director

__________ 20 ______ ______________________________

Dr André Roy, Dean

Faculty of Arts and Science

Abstract

Machines That Learn: Aesthetics of Adaptive Behaviors in Agent-based Art

Jean-Sébastien Senécal, Ph.D.

Concordia University, 2016

Since the post-war era, artists have been exploring the use of embodied, artificial agents. This artistic activity runs parallel to research in computer science, in domains such as Cybernetics, Artificial Intelligence and Artificial Life. This thesis offers an account of a particular facet of this broader work — namely, a study of the artistic practice of agent-based, adaptive computational artistic installations that make use of Machine Learning methods. Machine Learning is a sub-field of the computer science area of Artificial Intelligence that employs mathematical models to classify and make predictions based on data or experience rather than on logical rules.

These artworks that integrate Machine Learning into their structures raise a number of important questions: (1) What new forms of aesthetic experience do Machine Learning methods enable or make possible when utilized outside of their intended context, and are instead carried over into artistic works? (2) What characterizes the practice of using adaptive computational methods in agent-based artworks? And finally, (3) what kind of worldview are these works fostering?

To address these questions, I examine the history of Machine Learning in both art and science, illustrating how artists and engineers alike have made use of these methods historically. I also analyze the defining scientific characteristics of Machine Learning through a practitioner's lens, concretely articulating how properties of Machine Learning interplay in media artworks that behave and evolve in real time. I later develop a framework for understanding machine behaviors based on the morphological aspects of the temporal unfolding of agent behaviors as a tool for comprehending both adaptive and non-adaptive behaviors in works of art. Finally, I expose how adaptive technologies suggest a new worldview for art that accounts for the performative engagement of agents adapting to one another, which implies a certain way of losing control in the face of the indeterminacy and the unintelligibility of alien agencies and their behaviors.

À mes familles biologiques et étendues

To my biological and extended families

Acknowledgments

My first thanks go to my advisors: Chris Salter, who has been a constant source of motivation and challenge, pushing my reflections and practice beyond borders neither him or I would have ever foreseen; Jean Dubois, who has been my mentor since I moved into the arts a decade ago and has been generously feeding my reflections about contemporary art and new media ever since; and Bart Simon, whose openness doubled with pragmatism have widened my horizons as a researcher in the humanities.

I am also grateful to Yoshua Bengio who taught me the importance of intuition in science and introduced me to the fascinating field of Machine Learning; to Robin Dupuis, who first taught me about the importance of collaboration and experience in art; and to Lynn Hughes, whose generosity and passion for art, science, and technology had a deep impact on my path as an artist and humanities scholar. Much thanks to Peter Cariani, for his patience in reading my dissertation, and also for providing me with his insightful comments and revisions.

I would like to express my sincere gratitude to my friends and peers Aurélie Besson, Stéphane Couture, Alannah Fitzgerald, Anne Goldenberg, Nancy Labonté, Cécile Martin, Jaclyn Meloche, Karine Rathle, and Harry Smoak: our conversations have always been so fruitful and interesting, and your support and counsels have helped me survive through this endurance run. I would also like to thank my collaborators on the different research-creation projects that are described in this dissertation: Samuel St-Aubin and Stephen Kelly (*Vessels*); Marije Baaldman and Elio Bidinost (*n-Polytope*); as well as Joseph Malloch and Stephen Sinclair. Acknowledgements to the institutions that supported these projects: ComPeung, Avatar, Centre d'Art I Natura, Klondike Institute of

Art & Culture, LabXmodal, LABoral, the Center for Art Tapes, Perte de Signal, and Eastern Bloc. Finally, a special thank to Anke Burger, who fed me with green tea, vigorous optimism, and the most stimulating conversations.

Mes remerciements sincères à mes ami-es Marie Aboumrad (qui aime dire qu'un doctorat est plus durable qu'un mariage car il est impossible de divorcer de sa thèse), Patrice D'Amours (mon complice en matières de co-habitation et de discussions interdisciplinaires), Michèle Martin (qui m'a aidé à améliorer mes compétences d'écriture en anglais et m'a introduit à la pensée critique), Laura Noël (qui a su me fournir une bonne dose de calme et de sérénité dans les moments difficiles), ainsi qu'Adèle et feu Frenchie pour leur immortel amour nonhumain.

Enfin, j'aimerais remercier ma famille: mes parents Édithe Gaudet et Paul Senécal pour leur indéfectible soutien, leur amour infini, et pour m'avoir prêté à maintes reprises leur chalet à Morin-Heights pendant ma rédaction; et mon frère Étienne Gaudet ainsi que sa chienne Sally, pour m'avoir sorti de mes livres aux moments les plus opportuns. Thanks to Bob and Christel Gee, for feeding me vegetarian burgers and muffins while I was writing in their beautiful home in Regina; and, at last, my earnest thanks to Erin Gee, who encouraged me with inspiriting enthusiasm during my doctorate studies, offering me invaluable insights and help in proof-reading, feedback, and reviewing the final version of the text.

Thankful acknowledgement for funding to Fonds de Recherche du Québec - Société et Culture and Concordia University.

Preface

My fascination for nonhuman agencies goes back to my childhood. In the first seven years of my life, I was an only child, and I believe that I was uniquely good at it. Being very much so a calm and solitary kid, I did not need nor seek out friends of my age, as I seemed to get along better with adults and things. One of my favorite activities, besides reading books and fiddling with LEGO blocks, was to play board games. While the games I enjoyed most were meant to be played socially, I immensely preferred to play alone, and for that to work I needed to invent an opposing player, whom I would quite unimaginatively call “L’Autre” (“The Other”). “L’Autre” was a smart player of course — almost as smart as I was — and he often offered a challenging battle, although he always ended up losing, either because of bad luck or — when the die were rolling too much in his favor — due to some bad decisions he would make in the mid-game.

Fast-forward a few decades later. Upon entering the University of Montreal in my early twenties, I was quickly hooked by computer programming, and in particular by Object-Oriented Programming (OOP). During my first Java class, I started building a software library for making artificial beings as soon as I learned about class inheritance and packages, which was more of an abstract fantasy than anything and never really achieved much. In my second year, I read an interview about Machine Learning with Yoshua Bengio in the student departmental journal. I was immediately excited by the idea of computational processes able to make inferences, to come up with their own decisions, by interpreting real-life data. Machine Learning seemed to be a much more promising approach to creating Artificial Intelligence than the kind of rule-based logic I was seeing in my programming classes.

I started doing research with Yoshua during the Summer, and when I finished my B. Sc. I pursued a M. Sc. in his lab, studying neural networks applied to natural language modeling (Bengio and Senécal 2003; Bengio et al. 2006). I was especially enthralled by the spectacle of watching the error rate lowering during stochastic gradient descent. I imagined the neuronal connections self-adjusting as megabytes of English sentences from the Brown Corpus was fed into the system. I pictured the system as an artificial entity who tentatively tried to make sense of all of this text, exploring the millions of dimensions of the error space, self-organizing both semantic and syntactic information in a global, distributed network of subsymbolic representations.[1] I always perceived these systems as alive, in their own way, as possessing agency, years before I even know anything about Cybernetics and Artificial Life.

When I finished my degree and was offered a pretty sweet deal for pursuing doctoral studies in the area, I found myself in a difficult spot. I loved the work — at least, the creative part of it — but not the environment. The science itself was fascinating, but what seemed to come with it was less attractive: I perceived a very competitive and somewhat "macho" culture pervaded by rampant libertarianism, techno-utopianism, and a relative lack of interest in the philosophical and socioeconomic repercussions of the technologies we were developing. This, aligned with a generalized absence of self-scrutiny and self-criticism with respect to the entanglement of science with what were to me questionable endeavors such as financial computing or marketing which had alienated me since my first day in the lab, pushed me to quit the field.

I created my first art installation in 2004 in collaboration with fellow developers and artist Jonathan Villeneuve. The piece, which was created as part of a one-night multidisciplinary event for emerging artists, was an interactive video work where the audience's faces would get recognized through the use of a homemade implementation of a cutting-edge Machine Learning algorithm, and transformed using concave and convex distortions. It was really simple, quite silly in fact, but surprisingly efficient in drawing people into an interactive experience. What attracted me the

[1]The model I studied during my Master involved a Multi-Layer Perceptron feedforward neural network with eighty (80) hidden units, with millions of self-adjusting weights, which is relatively small by current standards. My Thesis focused on a technique for accelerating the training of such networks using importance sampling. Even with that acceleration method, in those years, it needed to run on a cluster of 20 CPUs for weeks in order to return results (Senécal 2003).

most to the experience was how the audience engaged with the work, using it in unexpected ways, exploiting its imperfections. That night, I decided I wanted to become an artist. In the coming years, I would, indeed, dedicate myself fully to this enterprise.

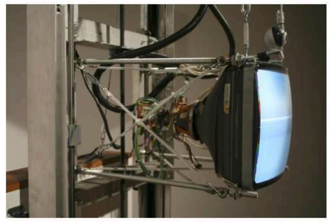
Changing disciplines is somewhat similar to moving into a new country. You need to learn new languages. You discover a lot about yourself through another culture's eyes. And yet, you never ever really feel at home anywhere anymore. An emerging artist in the mid–2000s, I was lucky to find in Montreal's blooming new media art scene an extremely welcoming and generous community.

At first I just wanted to put my knowledge in Machine Learning to use in my work, but this in turn constrained the possibilities for me to explore artistically, at a time where I needed, on the contrary, to "open up". So I decided to avoid using Machine Learning in my pieces, because I found that it limited my perspective as an artist. For more than five years, my brain had been trained to do science: now I needed to rewire it to be more of an "artist's brain", or maybe to develop another part of it. It was rather schizophrenic in a way, and it took me years to develop another way of thinking and being. Eventually these two realms inside my psyche, which were separated at first, came to reconnect in ways that are hard to explain, the result of which is what makes me feel constantly a bit like an outsider both in art and science — like an immigrant never feels at home even when traveling in his country of origin.

My early works consisted in artistic software-based works based on dynamic models of social and life processes. While interactive and computational in nature, these works stemmed from my original interests in adaptive intelligent systems and natural language processing, and were mostly an attempt to generate a poetic representation of some aspect of reality. My work with the Drone collective — a Montreal-based group of practictioners consisting of three programmers (Mathieu Guindon, Julien Keable, and myself), an electronics engineer (Samuel St-Aubin) and a visual artist (Jonathan Villeneuve) — was centered on dynamic sketches of social interactions, such as dialogue (*Vélodrame*, 2005) and tourism (*Travel Agent*, 2005). In the web-based software artwork *CHARACTERS* (2005–2006), a game based on the definition and evolution of real and fictitious identities is put forward within the constraining framework of an online dictionary.

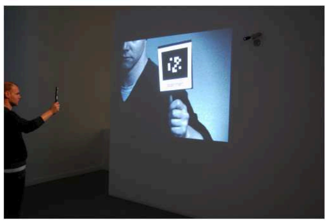
In 2005 I started a M. A. in Communication at Université du Québec à Montréal (UQÀM) in

(a) *Trace L* (2007). With Jonathan Villeneuve, in collaboration with Myriam Bessette. Photo by Alexis Bellavance.



(b) *Trace V* (2007). With Jonathan Villeneuve. Photo by Alexis Bellavance.

(c) *Trace S* (2008). With Jonathan Villeneuve. Photo by Alexis Bellavance.



(d) *Flag* (2008). Photo by Marleen van Wijngaarden.

ANNIELE `fem.` (French, 1969; from *annie(2)* and *madeleine*) CHRISTIANISM A married 39-year-old bisexual woman from Canada.

(e) *CHARACTERS* (2005–2006).

(f) *Vélodrame* (2005). With the Drone collective.

Figure 1: A sample of my early work (2005–2008).

the Interactive Media program of the School of Media, through which I produced a solo interactive installation (*Flag*, 2007) which confronts the spectator with an experience involving different mechanisms related to prejudice, dialogue and culture shock. Finally, the triptych *Trace* (2007–2008), realized in collaboration with artist Jonathan Villeneuve, yet again deals with depicting social and individual processes, such as attraction and memory.[2]

In the Spring of 2007, I went to Rotterdam to complete my M. A. in Communication studies. During that period, I participated in one of Piet Zwart Institute's "thematic projects" with artist and designer Kristina Andersen. The workshop, which took place through the whole three months of my stay in the Netherlands, dealt with physical computing in contexts of interaction design and artistic practice. There, I was confronted with new techniques and ways to think about interactivity and experience which brought me to create my first electronics-based artwork. The piece, called *Drift* (2007), was presented at the V2 Institute for Unstable Media as part of a group show and

[2]It is around those years that I started to use the *nom de plume* Sofian Audry to sign my artistic work.

was also my first work using reinforcement learning. It consisted of an old speaker attached to a handheld microphone. The interactive agent's only possible actions were to stay silent or emit a digital sound which was parameterized by a simple genetic algorithm, thus allowing an evolution of variety.

People could interact with the system by speaking into a microphone. The interactive agent would cycle through two different states, one where it would seek company and another where it would try to be alone. How to achieve these two goals was left to the agent to figure out. Typically, after a certain time, it would learn to attract spectators and respond to them using its "voice" when it wanted company, while staying silent when seeking solitude.

My experience with *Drift* would bring important changes to my artistic practice of the time. My contact with electronics and especially the Arduino platform opened up the possibility for a new approach to art through direct interventions in real-life using small, autonomous electronics objects rather than representing processes within the safe and artificial realm of the gallery walls.

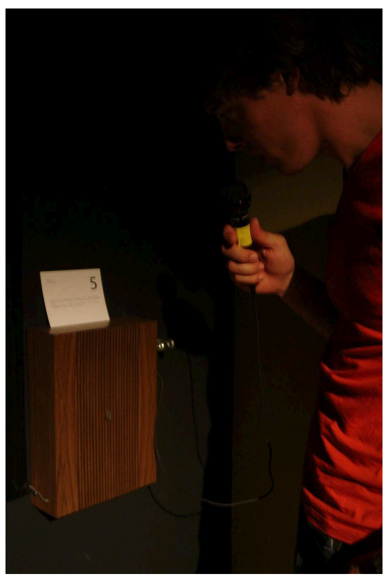


Figure 2: *Drift* (2007), V2 Institute for the Unstable Media, Rotterdam, Netherlands. Photo by Sofian Audry.

During the years to come, I would steadily move my practice towards what I call an agent-based practice, where in essence, my approach as an artist is to design an artificial agency. Here, “agent” is to be taken in its very general sense, of an entity or process that is able to act in an environment in response to its own perceptions. Moreover, these developments slowly brought back the idea of learning, which through the project *Absences* (described in chapter 4) would prove to be crucial component of my research in understanding the aesthetics of such agents.

Related to the concept of agent is that of behavior, which is understood here as the observable patterns that agents produce beyond their physical appearance. One can see how this idea directly resonates with *computationalism*, a philosophical view that understands cognitive processes as directly derived from algorithms (software) operated by the brain (hardware), and holds that human performance is completely independent of the material substrate that implements it (i.e., the body). But this concept of behavior need not be understood within this reductionist framework, and should rather be embraced by considering behaviors as the perceived performance of an embodied entity acting within its environment.

Computation is a central idea in new media, and could be perhaps the concept that distinguishes it from other artforms. In particular, computing allow to produce another artistic medium through the design of agents governed by algorithmic processes: an “aesthetics of behavior” (Penny 2000, 398). To make a parallel, if we compare video and photography and try to find the most fundamental properties that differentiates them, we can say that video, as a sequence of fixed images, adds to photography a third dimension: that of time. It does not mean that video is better than photography, but this difference is crucial to understand how both these media work, the effects they can create, and how to use them.

When it comes to computational behaviors, which are activated by computer algorithms, we are faced with something slightly different than video. A video is delimited by a finite time period: if you play it back, it will replay exactly the same sequence of images. A behavior is different: it can play for an infinite amount of time and will never exactly repeat itself. Yet, despite the inexhaustible nature of its manifestations, it is still recognizable by a human observer as a definite thing. If we experience a behavior long enough, we can adapt to it, we can get to know it, and then the patterns will become familiar.[3] Contrary to a mere record of agents' behavioral patterns (such as a video of birds flocking), a behavior can be affected by stimuli in real-time.

Agent-based artworks thus conjure "The Other", an alien entity that evokes liveliness, suggesting the emergence of novel aesthetic experiences. In seeking the creation of such experiences, I am especially interested in the blurry and muddy aspects of these behavioral forms, the uncanny nature that results from their imperfections, mirroring our own fragility as living beings. This might explain why I am an artist and not a scientist: I am not after some kind of optimal path to some objective truth, but rather, I believe in art's potential to provide humanity with the truth of reflexive becoming, the reflection of ourselves as imperfect and contradictory beings. Defect, faillibility, and indeterminacy are the substance of life, and the essence of freedom.

But beyond the aesthetics of behavior lay an important question: What role do algorithms themselves play when they are articulated through an embodied system? There is an important assumption as the base of this research: that there is a relationship between the choice of the algorithm governing an agent and the way this agent's behavior is experienced by the audience. This belief is supported by the fact that people are able to recognize behaviors beyond the appearance of the agents they animate. For example, people recognize swarming patterns in dots moving on a screen governed by a certain algorithm that *behaves* like a swarm (fig. 3).

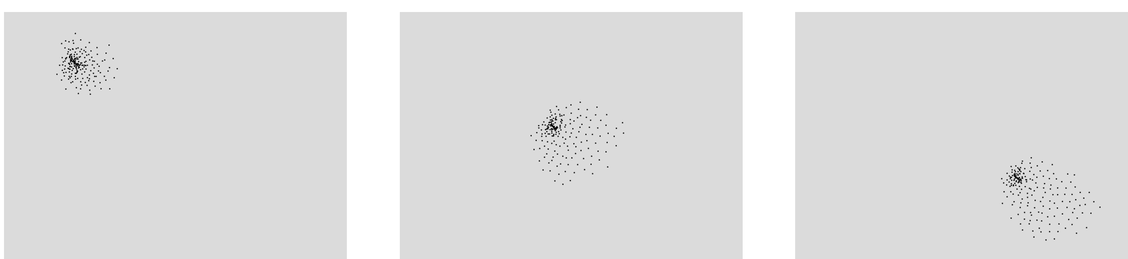


Figure 3: Swarming dots moving on the screen according to a swarming algorithm, as described in (Reynolds 1987). Based on code by Daniel Shiffman (Shiffman 2012).

[3] As a comparison, consider the behavior of speaking a natural language. While the possible combinations of words in a given grammatical structure in a given language are potentially infinite, in such a way that it is highly possible that the exact words that constitute this dissertation have never been written (fingers crossed!) and will never be written again in the history of the Universe, it remains that this thesis is recognizable as English writing, and can be understood as such by an English reader.

Does this mean there is always and automatically a relationship between an algorithm and how it is perceived? No. While two different algorithms will objectively yield different outcomes, it does not mean that these outcomes will be perceived differently by a human subject. For example, two algorithms might both generate different forms of noise, but they might be indistinguishable from one another by a general audience; the same way, two different "flocking" algorithms might give the same impression of "lifelikeness" to an audience without necessarily being discernable from one another.

Does it mean that we can control the outcomes by changing the algorithm? This is the assumption that this thesis makes, that it is indeed possible to do so, and that Machine Learning is a path for accomplishing this when working with self-organizing agents evolving in the real world.

At its heart, this dissertation focuses on the challenge of harnessing aesthetic experience from the building blocks of life, a story that keeps repeating itself, spanning different time scales in the history of humanity. It is a story of agents called genes who have traveled through millions of years, adjusting themselves, fighting and cooperating with others to survive through the organisms that host them; of an ant colony finding a route to a new source of food; or of an artificial neural network learning to pilot an automobile. It is the story of computationally evolved circuit boards that learn to recognize sound signals using microscopic magnetic perturbations of completely disconnected components, in ways that lay beyond human comprehension. This story is the story of all and everyone of us, that of a child learning how to smile, move, walk, talk, and later, ride a bike, read, make friends, step by step, through trial and error, exploring and exploiting its environment, set on a road of becoming. It is about how our world is filled with agents that adapt to one another, competing and collaborating in incommensurable ways that defy logical understanding. Exploring this universal narrative — and the evocative aesthetic potentials it holds — is what I seek above all in conducting in this research.

Contents

List of Figures

List of Tables

Glossary

adaptation

property of a system to adjust its structure in order to improve its performance in its environment

agent

autonomous entity that acts in an environment, usually in response to observations

artificial intelligence

branch of computer science that aims to reproduce intelligent behavior in computational systems

behavior

temporally invariant form of observable events produced by an agent performing in its environment

machine learning

sub-field of Artificial Intelligence that develops and uses computer programs that learn from data rather than relying on explicit logical rules

Acronyms

AI

Artificial Intelligence

AL, ALife

artificial life

ANN

Artificial Neural Networks

GA

genetic algorithm

GP

genetic programming

HCI

human-computer interaction

ML

Machine Learning

MLP

Multi-Layer Perceptron

PDP

parallel distributed processing

RL

Reinforcement Learning

SOM

self-organizing map

STS

science and technology studies

UX

user experience

Chapter 1

Introduction

> When viewed on a long enough time scale, life forms are always changing, adjusting, producing novel responses to unpredictable contingencies, adapting and evolving through blindly opportunistic natural selection.
>
> – Mark A. Bedau, *The Nature of Life*

> There is no intelligence where there is no change and no need of change.
>
> – H. G. Wells, *The Time Machine*

Since the 1960s, artists have been creating bodies of work using and/or inspired by computer technologies. In this research, I am interested in a specific branch of artistic works that make use of artificial agents, that is, man-made autonomous systems who act within their environment in response to what their perceptions. Examples include pioneering cybernetic artworks such as Nicholas Schöffer's *CYSP1* (1956) or Edward Ihnatowicz's *The Senster* (1970—1974); more recent works include Bill Vorn and Louis-Philippe Demers' large-scale robotic piece *La Cour des Miracles* (1997), Ken Rinaldo's artificial life installation *Autopoiesis* (2000) and Yves Amu Klein's "living sculptures". Artist and media theorist Simon Penny calls these kinds of work "embodied cultural agents" or "agents as artworks" and integrates them within the larger framework of an "aesthetic of behavior": a "new aesthetic field opened up by the possibility of cultural interaction with machine systems" (Penny 2000, 398). These works are distinct from so-called "generative art" or "algorithmic art" which use computer algorithms to produce stabilized morphologies such as images and sound. The former's aesthetics are about the performance of a program as it unfolds temporally in the world through a situated artificial body.

This project offers an account of a particular facet of this broader work — namely, a study of the artistic practice of agent-based, adaptive computational artistic installations that make use of Machine Learning methods. Examining the cultural-social-technical repercussions that arise in the use of such techniques in artistic works, I argue for an aesthetics of adaptive agents rooted in the distinctive way their behavior evolves and stabilizes as they couple with their environment.

Machine Learning is a sub-field of the computer science area of Artificial Intelligence. It employs mathematical models that can classify and make predictions based on data or experience rather than on logical rules. Learning systems usually consist of computational structures that adjust themselves when submitted to large quantities of data. Machine Learning is directly related to the biologically-rooted concept of adaptation which refers to a "process whereby a structure is progressively modified to give better performance in its environment" (Holland 1992, 7).

A late offspring of the cybernetic revolution, the field of Machine Learning has experienced an impressive growth since the mid–1990s. Its applications are extremely widespread and its success in the era of "big data" since the beginning of the millenium has pervaded Artificial intelligence research in areas such as pattern recognition, natural language processing, data mining, search algorithms and robotics.

As a proof of the importance of the field in contemporary society, consider the increased acquisition of Machine Learning startups by major IT players like Google, Facebook and others. For example, Geoffrey Hinton, the great-great-grandson of logician George Boole and an emeritus professor at University of Toronto in the field of Artificial Neural Networks, joined Google in 2013 as Distinguished researcher. Yann LeCun, another eminent pioneer in the field, was appointed in 2012 to be the first director of Facebook AI Research in New York City while Andrew Ng, an authority in the field of Reinforcement Learning, became Chief Scientist at Baidu Research in Silicon Valley in 2014. The year before, during Fall 2013, Ng's Machine Learning class at Stanford was the most popular class on campus, enrolling more than 760 students (Markoff 2013). His online class attracted more than 100,000 students in 2011, contributing to the development of Stanford's Massive Open Online Courses and the founding of Coursera (Friedman 2012).

Although the impact of these technologies on contemporary society is still relatively elusive, the debate surrounding them has reached the public sphere. In a recent appearance that made headlines, physicist Stephen Hawking warned about the threats that AI and in particular Machine Learning pose to the future of mankind. He signed an open letter asking for more control over AI and made numerous claims in the media that the rapid development of Machine Learning could allow us to reach human-level AI soon, possibly leading to human extinction.

Many researchers in the field have since refuted his arguments, showing that research was actually not progressing as dramatically fast as Hawking claimed and that humankind should not worry about an "AI singularity" happening in any foreseable future (Madrigal 2015). While Hawking's concerns seem largely unfounded, recent advances in Machine Learning research seem to have agitated, in Western media, the classic fear of machine cognition outstripping human intelligence, a sign that the widespread presence of AI is starting to gain public attention, for better or for worse.

While recent years have seen the field of Machine Learning grow at an unprecedented rate, the underlying idea of a computational system able to adapt to or learn from a flow of observations coming from real life is certainly not new. On the contrary, it recurs throughout the history of computing, from early concepts of negative feedback in Cybernetics to evolutionary computation. There has been a growing trend, since at least the late 1990s, to apply Machine Learning to the fields of robotics and agent-based systems (Dorigo and Colombetti 1997; Riedmiller and Merke 2002; Quinlan 2006; Chalup, Murch, and Quinlan 2007). Reinforcement Learning, a branch of Machine Learning dealing with agents adapting to their environment, is finally gaining momentum after more than a decade of research (Soni and Singh 2006; Gu and Hu 2002). The advent of Deep Learning in recent years suggest that this movement is not about to slow down (Mnih et al. 2013).

Despite this increased use of Machine Learning in many facets of contemporary industrial and commercial culture, one site where it has not seemed to make a meaningful impact is the field of art practice. This seems odd, considering that the use of computational systems in art goes back to at least until the early 1950s. Indeed, it feels like interest within the arts has been focused on techniques and concepts such as self-regulation, evolution and emergence, while there has been little rigorous work on Machine Learning and Adaptive Computation by artists (Kac 1997; Penny 1997; Tenhaaf 2000).

Why is this? The relatively recent popularization of Machine Learning in scientific communities might partly explain it. Another factor that may have slowed the adoption of such techniques by artists is the lack of access to the skills and knowledge required to utilize them. Moreover, there is a problem of translation: these techniques have never been designed for artistic production, which makes it particularly unclear for artists as to how they would even begin to use them in their own domain. Finally, the concepts surrounding adaptive systems and their definitions are fluid and shifting depending on the context in which they are used. For example, the definition of concepts such as "learning", "adaptation" and even "AI" as they are used by artists more than often differ largely from their scientific descriptions. The presence of such approaches in artistic works is often hard to trace because they are frequently used more as metaphors than as actual techniques.

Hence, while a significant and increasing number of new media artworks are indeed employing artificial agents, the vast majority of these agents are nonadaptive. Nevertheless, the integration of such methods in new media artworks raises important questions that have powerful sociotechnical and philosophical ramifications for aesthetic practice: (1) What new forms of aesthetic experience do Machine Learning methods enable or make possible when utilized outside of their intended context and are instead carried over into artistic works? (2) What characterizes the practice of using adaptive computational methods in agent-based artworks? (3) And finally, what kind of worldview are these works fostering?

These questions are hard to grasp, as they are solidly entangled with multiple disciplinary fields. The category of artworks under consideration here are built around situated agents engaged in an adaptive performance with their environment. But what exactly distinguishes them from systems that share all their characteristics but which are not adaptive?

One potential key to this question lies in the intimate relationship between adaptivity and performativity of a system. Adaptivity is what allows a system such as the brain to do things in the world. But the way a brain — and hence a human subject — performs is diffferent from the way a drop of water or a grain of sand does.

Cognitive scientist Stevan Harnad provides a comprehensive explanation of this concept when he talks about the importance of learning in category formation:

> The adaptiveness comes in with the real-time history. Autonomous, adaptive sensorimotor systems categorize when they respond differentially to different kinds of input, but the way to show that they are indeed adaptive systems — rather than just akin to very peculiar and complex configurations of sand that merely respond (and have always responded) differentially to different kinds of input in the way ordinary sand responds (and has always responded) to wind from different directions — is to show that at one time it was not so: that it did not always respond differentially as it does now. (Harnad 2005, 3)

I suggest that the use of Machine Learning in works of art is distinguished from nonadaptive works in their temporal unfolding. Works based on nonadaptive autonomous agents are, in theory, able to respond to interactions in real-time, often in complex ways. However, their behavior — understood as the way they act in the world — remains fixed over time, because the structure that implements their actions remains unchanged. The behavior of such nonadaptive agents can surely be convoluted and unpredictable, but more analogous to the way a grain of sand is carried by the wind (responding to inputs in a manner that remains fixed over time) than in the way a living and/or cognizing agent acts (which changes temporally as it is confronted to its environment).

In fact, adaptive agents can move beyond given limitations because their structure itself at a given moment is changeable in response to interaction with the world. An adaptive systems' behavior at any given time is determined by a structure, such as a set of weights in a artificial neural network or a digital DNA code in the case of genetic algorithms. When the agent acts in its environment, for example through its motor system, it does so both in reaction to both sensory data as well as its structural characteristics. Furthermore, this structural change is accomplished with the intent of enabling the agent to perform its tasks more effectively in the future. In other words, the history of the agent's interactions modifies the its behavior: the past feeds the future.

But the differences do not stop there. Indeed, one of the important characteristics of Machine

Learning algorithms (and especially the most recent advances) is their ability to represent the raw data in a more compact, abstract, efficient way (Bengio, Courville, and Vincent 2013). In other words, recalling Harnad, their ability to categorize. These systems accomplish this in their own unique way, thus the categories they create don't necessarily correspond to what we would expect as human beings. The responses from hidden neurons in a trained artificial neural network, for example, are often hard to grasp, if not utterly incomprehensible to a human observer.

Based on both observation and experience, I argue that the strong representational power of these systems, defined along dimensional axes that highlight the invariances in sensory data, can somehow be "felt" in the highly nonlinear behavioral patterns generated by these algorithms. This uncanny feeling feeds on the same intricate dynamics that evoke our very own way of performing into the world, nevertheless at the same time alien to us, as the behavior of such systems follows convoluted rules that lie beyond human comprehension.

This resonates with composer Iannis Xenakis' call for the application of scientific techniques by artists for the generation of new morphologies (Xenakis 1981a). For Xenakis, art is a "crystallization", a "materialization" of human intelligence, wherein art is fundamentally engaged in the same universal, deductive, and socio-cultural dynamics on which the sciences are founded. As a manifestation of this claim, he notes the close historical ties between music and mathematics, demonstrating how one cannot be dissociated from the other. Xenakis concludes that a new type of artist is required, one who can freely use science and mathematics to create spatio-temporal "shapes" that can only be understood as the constant interaction between function and structure.

What does this mean for artmaking in a contemporary moment of Machine Learning and AI? This question is challenging to address, because the practice of adaptive agent-based artworks is marked, first and foremost, by a high degree of diversity in materials, subjects and outcomes. In order to analyze the practical aspects of artworks that integrate these technologies into their structures, I first need to take into account the technical challenges inherent to Machine Learning systems in the context of agent-based artistic installations. Namely, the difficulty to build big and reliable data in artistic venues, the loss of locus of control by the artist, difficulties that come with real-time adaptation and the lack of a descriptive framework for adaptive behaviors.

An important feature of this genre of work lies in the constant tension that exists between the scientific and the artistic perspective in the creative process. I concur with digital artist Marc Downie on the importance of authorship in this matter: it is a common mistake to think that the use of computational technologies could replace altogether the artist's artistic input, and it is crucial for artists to hold on to their aesthetic intentions (Downie 2005). This might be especially true for Machine Learning technologies which largely consist of optimization techniques that were never designed for artistic use.

The specific shape of adaptive behaviors that I have called attention to is not a magical trick but rather a tool to be explored to achieve a certain effect. Artists need to reflect from the beginning on how the overall experience of a work will be related to the adaptive process, and in general, to adopt a critical stance, in relationship to the technology.

It is useful here to stress that the specific aesthetic qualities of adaptive agents outlined above are almost haphazard, in the sense that scientists working in the field of Machine Learning typically have very little interest in the shape of behaviors and how they unfold in time. Most Machine Learning algorithms run entirely offline, training intricate mathematical models on huge, pre-compiled databases of real-world data, with the sole objective of achieving a better performance on solving a specific problem, according to a precise error factor. Bearing this scientific perspective in mind, the aesthetics of such processes appears to be little more than a side-effect.

In order to tap into the artistic potential of learning and adaptive systems, artists need to somehow invert this perspective. Whereas scientists fine-tune their algorithms to achieve better performance over an agreed-upon error measurement, artists need to find their own way through the different components of learning systems in order to produce subjectively compelling behaviors.

This being said, the artistic and scientific practices of adaptive agents have a similar set of relationships between author and machine in that they both involve a constant interaction between the practitioners and the material agents they interact with. Sociologist of science Andrew Pickering has come up with the concept of a "dance of agencies" to describe the constant movement of resistance and accommodations going on in scientific practice (Pickering 1995). The stakes in art are different than those in science, of course, because artists are generally invested in the creation of an experience, while scientists try to discover or confirm some truth about the world by building theories based on observations. However, both the artistic and the scientific processes have striking similarities, and what Pickering essentially says about scientific practice can be applied to art.

Bringing British Cybernetics to the foreground as an inspirational example of this worldview, Pickering argues for a conception of human cognition that is performative rather than representational: “the cybernetic brain was not representational but performative, as I shall say, and its role in performance was adaptation.” (Pickering 2010, 6). In making these statements, Pickering stands alongside many other humanities scholars who reject representationalism in favor of performativity (Hayles 1999; Penny 2000). Following my discussion on adaptive systems, I want to connect Pickering’s performative ontology of science to both Xenakis’ and John Cage’s approaches to indeterminacy in art.

As an artist and researcher trained in the field of Machine Learning, I propose to tackle these seemingly abstract and hard questions using three complementary approaches. First, I examine historical accounts of machinic life and machine intelligence since the post-war era from both the perspectives of computer scientists working in the field of Machine Learning as well as techno-cultural studies scholars exploring the larger sociotechnical impact of machine-based systems. Through this dual perspective, I touch upon issues of adaptivity, learning, autonomy, self-organization and emergence.

Second, I analyze artistic works making use of Machine Learning algorithms through close readings of core texts on adaptive systems in three areas: science and technology studies (STS) (Hayles 1999; Pickering 1995, 2010); media art history (Shanken 2002, 2015; Ascott 2003b; Whitelaw 2004) and computer science (Sutton and Barto 1998; Bishop 1995; Langton 1990; Langton 1995). My aim here is to tease out the two different worldviews I described earlier (representational versus performative) and to articulate how these different viewpoints have come to be defined, problematized, expressed and legitimated in artistic works utilzing computationally adaptive techniques.

Finally, I provide descriptive and reflexive accounts of practice on three works I have been involved in over the past few years. These artistic works specifically employ Machine Learning methods such as Reinforcement Learning and Artificial Neural Networks in order to achieve the kind of adaptive behavior I have been describing above. I describe these works — *Absences* (2008—2011), *Vessels* (2010—2015), and *N-Polytope* (2012) — in detail because they each embody different ways of working with such systems to achieve certain aesthetic effects. Furthermore, they embody the tensions between representation and performance that I'm attempting to describe here in practice. As such, these artistic works are important to my argument in that they involve different ways to approach my research questions, grounding theory into real-life bodies of work.

1.1 About my Artistic Practice

This dissertation, due to its interdisciplinary nature, might seem at times very technical to the reader. Since I am writing this thesis primarily as an artist and a humanities scholar (and not as a scientist or engineer), I believe it is important to give an overview of my own perspective on art, and my approach as a practicing artist before entering into the core of the subject.

When it comes to art theory, I consider myself an anti-essentialist (Weitz 1956). It is never possible to "pin down" art, to find a common set of properties that would encompass all of its different forms. I think art is best conceived of as a socially constructed, constantly fluctuating concept marked by an incommensurate richness and diversity. Art can be recognized, rejected, criticized, debated, but never reduced to an absolute set of sufficient characteristics.

Whether political, expressive, and/or conceptual, however, most works of art propose a form of human experience that happens in context. In this sense, art as John Dewey argues, is chiefly about experience (Dewey 1959). It consists of physical energy and matter that circulate through human bodies, stimulating neural synapses, provoking hormonal reactions, mobilizing organic systems. At its best, art can truly change someone, physically, in ways that are often so personal that they can hardly be foreseen. This transformative interweaving between the artistic process and the perceiver is core to the aesthetic act.

The creative process itself is central to my practice. Part of being an artist is to be able to bend oneself according to the transforming materiality of the work. The work itself is thus in a process of becoming–through my interactions with matter (the materiality of the artwork), it grows an identity of its own. My engagement with practice does not only concern the interaction with these material agencies. A significant aspect of my art practice involves collaborative work, which necessitates an open approach. One needs to agree on an abstract set of basic principles, and then, each co-author usually advances the project with the skills s/he has, in constant feedback with the rest of the group.

This collaborative aspect of my artistic practice is directly linked with the nature of my preferred medium: computer programming. The immense power of computers does not lie so much in their capacity to rapidly treat information than in their great flexibility. While I believe the concept of a "universal machine" is problematic in many ways — some of which will be covered in this thesis — one cannot deny that computers, as a technology, possess a unique capacity to adapt to different contexts and situations, which explains to a large extent their widespread dissemination across all spheres of society.

This "quasi-universality" of computers makes them especially appropriate for collective work, because in themselves algorithms are somehow an empty shell. Only through their embedding in a network of other media can they truly become effective in the world. This brings me to mention another equally important quality of computation beyond their flexibility, which is their capacity to express behavioral patterns — in other words, to enact agency.

While computation has allowed the expansion of existing media such as photography, video, and music, more interesting to me is the new forms of media that it can generate. The temporal unfolding of dynamic patterns enabled by computation reveals movements that never exactly repeat themselves, yet can be experienced in real-time as "something" that can be "felt". In this regard, I have a particular interest in artificial intelligence, which is the field of computer science that has engaged primarily with questions of computational agency and behavior. As some of the core results of this dissertation suggest, one of the aspects of this question is that some algorithms exist that do not have a fixed structure, but can evolve over time.

From an aesthetic point of view, what are the politics of such practice? My own refusal of the technoscientific trajectory that was offered to me as an engineer, in order to pursue the risky venture of an artistic career is a deliberate political act. Nowadays, art provides one of the last remaining bastions for research against techno-utopianism and techno-determinism. I have no doubt that whatever companies engaged in AI such as Google and Facebook are after will "work", that we will be besieged by autonomous cars, auto-diagnosis health systems, and robots that perfectly reproduce human behavior. But the real challenge of history is not about technological advancement: it is about a process of becoming, it is about the kind of human qualities we want to develop as a species.

Science and engineering have, over the past 30 years, been caught in the turmoil of a technocratic, applied agenda, and have come to a dead-end. As an interdisciplinary artist trained in science and engineering, I want to use my unique position, and through my work, actively participate in the critique of these technologies, suggesting alternatives to this limited future.

1.2 Scope and Relevance

This research concerns agent-based installation artworks that use Machine Learning or adaptive computational systems as a core element of behavior generation. Works that use Machine Learning techniques without specifically staging agents, or nonadaptive agent-based installations, are also considered when necessary, but only in order to better grasp the concepts under scrutiny.

The relevance of this research project, highlighted by the originality of its approach and the importance of its subject of inquiry, can be summarized as follows:

1. Adaptation and learning are important concepts to understand the world we live in and the future of contemporary societies.
2. In particular, they provide core insights into sociotechnical questions of practice in both art and science.
3. Art offers a way to critically engage with adaptive systems through their material articulations, in a manner that neither science or the humanities can approach them, thus generating alternative kinds of knowledge.
4. However, while there has been some work on related questions of emergence and interactivity, there is currently a lack of aesthetic theories specific to adaptive systems and how they are,

have been, and can be used, in artistic practice.

This research offers an interdisciplinary account of adaptivity and learning in machinic agents. I show how adaptivity pervades contemporary conceptions about life, autonomy, cognition, intelligence, and the brain, emphasizing its strong influence on media art since the 1960s. Underlying the concept of adaptivity is the idea that the human brain should not be understood as a universal machine for solving problems using logical rules, but rather as an incredibly malleable organ with the ability to change, to tune itself to its environment, and to even reinvent itself when needed. Adaptive artworks resonate with this idea, holding the potential to change our perception of the world, a world filled with performative agents in constant flux, adapting to their environment and to one another.

This study is *not* about generative art and design. I am not interested here in computational algorithms that create fixed and stable forms, but rather, I am interested in the aesthetics of the processes themselves as seen in their real-time unfolding, and as part of an embodied, material experience; agents that live and act in the physical world. As there has been much research carried out in adaptive music composition and improvisation, I focus the scope of my research outside of the music realm.[1]

Because I am interested in considering adaptive systems as an art practice, I choose not to engage with works that merely make use of Machine Learning techniques as part of a specialized pattern recognition component, such as face-tracking devices, unless when this component is modified or used in a critical fashion that influences the aesthetic behavior of the piece.[2] Finally, further limiting the scope of analysis, I will not be considering works that involve human performers, such as theater or dance works. Instead, I will focus on experiences that involve a direct, physical relationship between nonhuman adaptive systems and a human audience.

Finally, I feel it is important to note that despite the fact that I directly engage with these techniques as an integral part of an artistic practice involving the conception, design and implementation

[1] With the exception of certain robotic artworks that involve music performance such as Baginsky's artificial adaptive robotic rock band *The Three Sirens* (1992—2005).

[2] In other words, I exclude from this study artworks that use ML techniques in ways they are usually intended to be used.

of computational systems, I am not seeking advancements in the field of Artificial Intelligence, but rather, to connect art practices that make use of AI with broader questions of life and agency. The project thus remains within the field of the humanities, however, my hope is that it will provide contexts and approaches that could benefit AI researchers as well.[3]

1.2.1 Adaptation Matters

As necessary conditions of life, standing right between self-organization, autonomy, and the generation of novelty, adaptation and learning are powerful concepts for understanding living systems and they way they operate. Ultimately, it is the capacity to adapt that distinguishes life from other natural phenomenon. Life is about maintaining and extending itself in a changing environment, about learning from experience; it is a process of becoming that emerges through a constant negotiation with inner and outer conditions.

Adaptation and learning occupy a sweet spot in the hierarchy of systems properties, being more closely related to the living than the concepts of emergence and self-organization imply. Adaptation is a sufficient yet unnecessary condition of emergence, and while there exist self-organizing systems that are nonadaptive, such as hurricanes and galaxies, all adaptive systems must have a capacity to self-organize. I claim that adaptation is, in fact, the process by which living systems self-organize: as all living systems are complex, emergent systems, they are also adaptive, in their capacity to adjust their own structure and behavior to their environment.

Yet, few researches have addressed the questions of adaptation and learning in the fields of media art theory, art history and science and technology studies, let alone their relationship with concepts such as emergence, self-organization, self-regulation, autonomy, and life. Recent studies about "artificial-life art" or "behavior aesthetics" have mostly focused on concepts of embodiment (Penny 1997; Dourish 2001; Bogart and Pasquier 2013), emergence (Baljko and Tenhaaf 2008; Soler-Adillon 2015), and the generation of novelty in lifelike agents (Whitelaw 2004; Cariani 2008; Boden 2009). While these works are key to understanding the way natural and social systems operate, and

[3] For example, there is a growing community of AI scientists interested in notions of computational creativity and on artistic applications of AI, who may find in this study valuable criticisms and alternatives to the research trends commonly found in their field.

correspondingly what this might mean for the arts, they are missing an important piece of the puzzle, which this study aims to address.

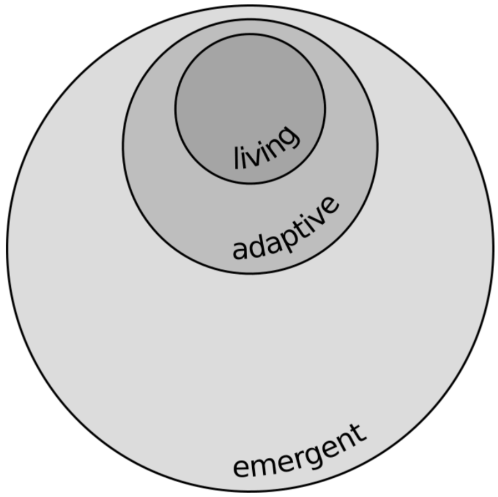


Figure 4: Hierarchy of systems. Living systems are adaptive (thought not all adaptive systems are alive). Adaptive systems are emergent (though not all emergent phenomena are adaptive).

The importance of adaptation and learning in contemporary views of natural systems is echoed by the strong and growing presence of these concepts in science particularly in AI technologies that we use every day. Since its inception in the post-war era, computer science has oscillated between two poles: one that sees life as a logic-based phenomena, and another which emphasizes the importance of adaptation. In the beginning of the 21st century, we are seeing the demise of the former and the triumph of the latter.

As deep neural architectures computing billions of artificial synaptic connections on GPU clusters owned by the largest IT companies, attuning to the smooth melody of our everyday actions in the most peaceful, steady, and inexorable fashion, the digital world we were used to, with its recognizable, explainable, decision-making procedures based on hand-coded heuristics, is already gone. We are moving into a new era, where pervasive, organic-like apparatuses feeding on statistics are replacing rule-based systems, adaptively coupling to our bodies in over-encompassing, distributed processes of control and optimization. To understand this new age, we need to extricate ourselves from an outdated vision of computational systems as formal, rule-based, logical constructs, and start seeing them for the biologically-inspired, statistically-driven, agent-based entities they have become.

This become particularly important because we now have technologies that are adaptive, whereas these kinds of systems were only found in natural phenomena before. In particular, such systems are newly important for artists and art theory because: (1) they suggest new approaches to work with emergent systems; (2) they hold the promise of generating more "lifelike" behavioral patterns, opening up novel ways to understand what it means to be alive and human; (3) they challenge the notion that artistic creation is a purely human-centric practice, as the agency becomes diffused between humans and machines that couple with one another.

1.2.2 A Core Dimension of Art and Science Practice

Adaptation and learning are key concepts by which to address the question of practice. Sociologist of science Andrew Pickering suggests the "mangle of practice", an ongoing dialectic of *resistance* (by nonhuman entities) and *accommodation* (by scientists), as a framework to examine scientific practice (Pickering 1995). Pickering describes the way human and nonhuman agents interplay in "the mangle" as a "dance of agencies". Science is thus best described as a performative material practice that stages both human and nonhuman agents, the former adapting to the latter.

More recently, Pickering finds in Cybernetics — an interdisciplinary field started in the 1940s which exists today under "many other names" — the perfect embodiment of his theory (Pickering 2010, 15). As early as the 1950s, cyberneticians built lifelike devices as a means to attain a higher understanding of the workings of the brain. For example, Grey Walter designed a pair of artificial "tortoises" with some basic learning capabilities, while Ross Ashby created the homeostat, a self-regulating system that aimed to mimic feedback processes in the human brain. While clearly scientific in nature, the creative process that made these apparatus possible is very close to art practice, at least in the domain of computation art.[4]

Another interesting insight comes from Greek polymath Iannis Xenakis. For Xenakis, art and

[4] In the 1960s and 1970s, some of these devices were shown as part of art exhibitions, such as Gordon Pask's *Colloquy of Mobiles* (1968), a Cybernetics installation that was presented in the "Cybernetics Serendipity" 1968 exhibition curated by Jasia Reichardt in London.

science can both access forms of objective knowledge through the processes of *inference* (the process of drawing ideas out of observation and reasoning) and *experimentation* (verification of these ideas through experiments). But art can go beyond these, attaining forms of subjective knowledge through what Xenakis calls *revelation*, giving us access to the emotional, personal, universal dimension of reality. In Xenakis' mind, the artist must thus be "simultaneously rational (inferential), technical (experimental) and talented (revelatory)" (Xenakis and Messiaen 1994, 5–6).

In inference, one looks at data and makes an hypothesis about the process that generated it; in experimentation, one verifies if the hypothesis is right, generating new data in the process. This is an adaptive process, where we iteratively refine our view based on our actions in the world and their consequential effect on our sensory surfaces. How does revelation interoperate with this process? Is revelation reserved, as Xenakis claims, only to the arts, or does it also appear in science, but hermetically, available only to scientists themselves?

Xenakis suggests that in the future, science and art must learn to work together, forming "alloys" with new, emerging properties. But Xenakis definitely takes sides. In his mind, art subdues science: it should be the driving force, bringing problems for science to resolve. I reject this asymmetric worldview and I suggest, instead, one that supposes a co-adaptive relationship between the artistic and the scientific spheres. Can adaptation help us understand and possibly establish such a relationship between art and science, one that goes beyond Xenakis' philosophy of art and science alloys?

1.2.3 Art Offers Alternative Ways to Engage with Adaptive Technologies

The coming-of-age of Machine Learning seems to be activating a mix of fear and excitement, turning contemporary discourse about AI technologies into a highly polarized debate. The first camp warns against the emergence of a much dreaded technological "singularity" from which point AI will replace humans as the superior intelligent species, with possibly dire consequences that could lead to the extinction of the human race (Kurzweil 2006). On the other side, Silicon Valley's techno-optimist choirs are chanting the libertarian utopia of a post-work, post-democratic world where all of humanity's problems are to be smoothly solved by benevolent artificial learning agents.

With their capacity to work both critically and creatively with material and experiential questions, artists have a unique standpoint for reflecting on the complex issues surrounding AI. Art can suggest alternative ways of engaging with AI systems and imagining our relationship with them now and in the future.

I claim that learning and adaptive systems suggest a complete change in paradigm about our way of considering technology and how it operates in the world. Technologies of the past and present are immensely nonadaptive: they are driven by a human-centric ethics that seeks to control nature (Pickering 2010). Quite paradoxically, current-day Machine Learning has not really escaped that paradigm, being used for the most part for pattern recognition purposes in attempts to efficiently solve concrete, measurable "problems"; to gain more control over outcomes.

What if technologies were designed to adapt themselves to natural processes and entities, rather than the other way around? Can we envision technologies that are not meant to control nature, but rather to take part in an ecosystem, trying to survive while allowing other processes to flow? Can we give artificial agencies the right to make mistakes? Can we allow them to be gracefully weak, imprecise and hesitant, just as we are? In the field of AI, what would happen if we moved beyond the ideal of optimization and control, towards the most open-ended paradigm of adaptation as a living process?

I believe adaptive systems allow us to imagine a whole new future for the world we live in. In that future, artificial agents would become an active part of the aesthetic fabric that makes up our existence. I picture adaptive agents acting as surrogates, carrying emotions in their neuroplastic shells, facilitating their contagion like viruses. Some would have their own survival attached to something or someone we hold dear, helping us protecting them. Some would write with us, dance with us, do things *with* us rather than *for* us — or, as STS scholar Sherry Turkle says, do things *to* us by "by changing the way we perceive ourselves and our sociotechnical environment" (Turkle 2006, 1).

New artforms will likely emerge beyond the traditional formats. Public works could run for long time spans, evolving across many generations, constantly adapting to new circumstances. Artificial beings could live inside homes, keeping a trace of past interactions in the way they behave and act in the world, transcending time. Robotic bands and human-machine collectives would emerge, live and die, producing albums and shows, breaking up and reassembling. Nomadic agents could be allowed to circulate among us, moving at the speed of light, influencing one another and engaging with us, stimulating debates around art, politics, and science, rather than merely providing us with what we are assumed to need. AI travelers could be sent in space, surviving for millions of years, able to adapt, grow, and die beautifully.

The development of Machine Learning has long moved away from Cybernetics, which was largely concerned with adaptive processes as a way to understand the living. This eventually resulted in the exploitation of massive amounts of data for optimization, classification, and recognition tasks. The vast majority of learning algorithms are designed to learn offline (i.e., not in real-time) in order to perform a given task. The few artists who use Machine Learning usually stay within the scope of their intended use[5].

The success of Machine Learning goes hand-in-hand with "big data", large collections of information which are, for the most part, in the hands of big businesses such as Google and Facebook. These companies generate wealth and power by appropriating these massive datasets which, while provided by the general population, stay out of the control of the public. These learning algorithms are increasingly present in our lives, often without us knowing. We do not see them, we do not understand them, and this leaves us ineffective at criticizing them or critically engage with them. We are left without a voice, to be the passive containers that corporate interests feed upon, for the benefit of private interests.

As Machine Learning algorithms continue to transform our world, it is crucial to develop alternative ways to approach these technologies beyond science and business. What I propose here is one attempt to do just that, by bringing together sociotechnical, artistic and aesthetic questions into a global framework, and by suggesting ways artists can manipulate these algorithms. My hope is that by providing these tools, I will inspire new ways of understanding the technology and its impact on our world, giving artists some agency in creating works of art that are free of corporate

[5]For example, see David Rokeby's works using ML for computer vision and image pattern recognition. In such cases, ML could be replaced by any other technique which would be more or less as efficient, without affecting the fundamental artistic intent behind the work.

power and constraints.

1.2.4 Adaptation in New Media Art Theory: Filling the Gap

New media art as a field of research has not often been a sustained topic of study for art historians, leaving a void that is only starting to be addressed. For a large part, it is new media artists themselves who have started building some of the theoretical tools for understanding their discipline through analyzing their own practices.

Still, theory about new media remains scarce, in particular when it comes to the niche of agent-based art, let alone that of adaptive behaviors. A search in the database of Leonardo (the most important peer-reviewed journal in the field of technological art) from 1997 to 2015 reveals a huge gap between the number of papers containing references to Artificial Life (168), Artificial Intelligence (160), and Cybernetics (165), when compared to Machine Learning (23), Connectionism (12), and Adaptive Systems (4).[6]

Table 1: Number of articles containing a reference to certain terms in Leonardo (1997–2015).

Expression	Number of publications
Interactivity	866
Artificial Life	168
Cybernetics	165
Artificial Intelligence	160
Self-organization	107
Machine Learning	23
Connectionism	12
Adaptive System	4[7]

By comparison, Machine Learning and Adaptive Computation have been an essential part of the AI ecology. While their role has often been peripheral, their presence has been exponentially growing since the Deep Learning revolution of the mid–2000s, largely due to their unprecedented success in tackling major AI-related problems. Adaptation and learning are thus critical concepts whose increasing presence in our world has vast sociotechnical repercussions.

[6]These results were compiled by performing a clear text search using the JSTOR Arts & Science search engine on August 3, 2016.

As these technologies become more popular and more readily available, their use in works of art will likely grow as well. Yet there exists at the moment almost no guidelines, tools, or theoretical frameworks on how to make these works and think about them. It is therefore crucial for the discipline of new media arts that we start building an aesthetic theory of adaptation, in order to allow for a better understanding of artworks that utilize them, as well as to understand the processes entailed in working with them artistically.

1.3 Contribution

This research project aims to address three interrelated questions about adaptive systems in computer arts: one is about aesthetic experience, one is about practice, and one is about the forms of knowledge with emerge from such contexts.

The first question concerns the kinds of experiences that are specifically enabled by using Machine Learning and Adaptive Computation as part of agent-based artworks. By extending and refining Simon Penny's behavior aesthetics (Penny 2000; Kim and Galvin 2012), I show that emergent and adaptive processes exist in a different kind of time than formal/nonemergent/nonadaptive behaviors, because emergent/adaptive behaviors change their morphology through time and thus bring with them the potential for spectators to experience the unfolding of time in novel ways. Therefore, these systems as used in art bring with them the potential to experience time in novel ways.

As introduced above, adaptive behaviors bring with them a sense of aliveness, because adaptation is "one step closer to life" than emergence. For instance, there exists emerging phenomenon that are nonadaptive, yet adaptivity is impossible without emergence. Adaptation is a necessary condition of life, first at the level of species development and survival (genetic adaptation) and second, at the level of the individual (neuroplastic adaptation). This "lifelikeness" comes with its own experiential essence, a quality which has been sought by artists since the dawn of time, from the animal representations in the caves of Lascaux to current-day agent-based installations.

The second question I investigate is how adaptive computational methods affect practice in agent-based artworks. I delve into the core dimensions that define Machine Learning algorithms: (1) the *task* they are aimed to accomplish; (2) the *model* that is trained by applying (3) an *optimization procedure* that uses (4) an *evaluation function* to measure the system's performance over (5) a set of *data*. In the process, I suggest ways these components can be exploited for artistic expression.

Connecting Cariani's taxonomy of agents (Cariani 1989; Cariani 2012) with Penny and Soler-Adillon's work on self-organizing systems (Penny 2009; Soler-Adillon 2015), I present a framework for understanding behaviors based on the temporal unfolding of their morphology. At the first level, one finds patterns generated by stateless, function-like devices called "mappings". The second category comprises behaviorful devices with states, driven by Finite State Machines or other formal structures that do not evolve through time, but generate recognizable temporal patterns. Finally, the third and last level is occupied by "metabehaviors", that is, behaviors whose transformation in time is driven itself by a behavior, such as those generated by adaptive or evolutive devices.

As adaptation is intimately connected to emergence, being the *way* by which self-organizing systems mutate their behavior in response to changes in their environment, I argue that Machine Learning and Adaptive Computation can provide a suitable tool, a pathway to design emergent behaviors that move beyond the direct — and often strenuous — programming of unitary agents.

My third and last contribution lies in the delineation of another worldview brought forward by adaptation in general, and by adaptive works of art in particular. Adaptation allows to imagine embodied artistic works that couple with the world, actively changing it. The integration of these systems in artworks challenges the way art is presented and received by audiences, as their lifelike properties also make them as intricate and mysterious as life is. Perhaps more than other media art forms, artworks that integrate adaptive systems demand more effort, more attention, as the public itself needs to be engaged in an adaptive endeavor. These works thus often demands that the audience spends enough time with these agents to get to know them in an embodied manner.

Through their capacity to transform their behavior through time, to reinvent their way of acting in a constantly changing environment, adaptive agents-as-artworks can allow the emergence of a worldview wherein agents are not just generating novelty out of the blue, but rather in relationship with one another, by tentatively pointing their behavior towards a constantly evolving environment.

1.4 Literature Review

This thesis aims to synthesize different perspectives in an effort to get a broad understanding of the notion of adaptivity and its evolution in contemporary discourse in art and humanities. This is no easy task, since different disciplinary fields usually evolve their own vocabularies and concepts. A term used in two different disciplines might mean completely different things, whereas two different terms might actually refer to a common notion. In practice, things are usually much more blurry and one needs to be extremely careful where to draw the line when trying to establish appropriate connections and groupings.

As previously argued, humanities scholars and artists who have tackled these concepts have often provided incomplete, confused, inaccurate and/or out-of-date accounts of these practices. An important contribution will thus be to use my scientific training in Machine Learning, mathematics and computer science to articulate, disentangle and update these accounts.

In bringing together scientific, artistic, philosophical and sociotechnical contexts, I hope to (1) present a more scientifically rigorous account than has previously been accomplished in media art history and STS; (2) show how the aesthetic questions that drive the thesis are intimately entangled with the scientific histories of AI, connectionism and adaptive systems, and; (3) to provide artists working (or wanting to work) with adaptive technologies with a set of anchorpoints, highlighting the aesthetic properties and potentialities specific to adaptive systems, and the challenges of using them in artistic contexts.

In order to address the complex issues teased out above, I thus draw on three distinct but overlapping bodies of literature and practice, namely: (1) scientific literature in computer science and robotics; (2) sociotechnical and philosophical perspectives of artificial intelligence and artificial life systems; (3) media art history and theory, including writings from art critics and media theorists as well as works of media art dealing with adaptive systems.

First of all, I examine literature in computer science, tracing through the history of Machine Learning since the post-war era. Within this history, I trace the occurrence and influence of adaptive systems and Machine Learning on the field of Artificial Intelligence, with a focus on connectionist systems. I focus not only on the sociohistorical context but also directly on the techniques themselves. The reason for looking directly at the technological practices is to address accounts of AI by cultural critics and theorists who rarely have a direct experience with the technology and/or provide inaccurate descriptions of these technologies (the same way many scientists writing about art usually only superficially address artistic questions). As an artist trained in computer science and AI, I believe it is important to go back to these techniques directly to come up with my own historical account of these practices.

In order to trace this archaeology of Machine Learning, I inspect seminal works in Cybernetics (Wiener 1961; Ashby 1957; Rosenblueth, Wiener, and Bigelow 1943), information theory (Shannon 1948) and early connectionist models (McCulloch and Pitts 1943; Rosenblatt 1957; Selfridge 1959). I consider the emergence of what is referred to as "classic AI" or "Good Old-Fashioned AI" (GOFAI) in the 1950s and 1960s, marked by a strong optimism in the ability of purely symbolic, disembodied computational systems — often referred to as "computationalism" — to achieve human-level cognition, and how this politically and instititutionally lead to the abandonment of the connectionist project (Newell, Shaw, and Simon 1959; Minsky and Papert 1969). I describe the emergence of the field of Artificial Life (ALife) which brought together ideas on self-replication, self-organization and emergence, supported by a "bottom-up" approach and a computationalist definition of living systems (von Neumann 1951, 1966; Langton 1986, 1989b; Ray 1991; Reynolds 1987).

In parallel, I explore the emergence of Machine Learning in the mid–1980s, following the demise of symbolic AI and marking the end of the "AI Winter", by focusing on approaches in neural computation and pattern recognition (Rumelhart, Hinton, and Williams 1986; Bishop 1995; Duda, Hart, and Stork 2001), genetic algorithms (Holland 1992; Mitchell 1998) and Reinforcement Learning (Sutton and Barto 1998; Wiering and Otterlo 2012). I contrast Machine Learning to "Nouvelle AI", an approach to AI which suggests that AI systems should be built incrementally, starting with "simple levels of intelligence" that do not use any representation, and rather, "use the world as its own model" (Brooks 1987). A major criticism of Nouvelle AI in the field of robotics is that by refusing to include any form of representation in its architecture, it throws away Machine Learning, thus making it difficult to engineer adaptive systems which are deemed necessary to achieve more robust forms of intelligent responses to real-world problems, such as driving a vehicle (Ziemke 1999).

For example, an important research strand since the 1990s in the field of robotics has been working with Machine Learning methods, often used in conjunction with architectures and algorithms inherited from Nouvelle AI or symbolic AI (Dorigo and Colombetti 1997). Current research seems to advocate for hybrid approaches that integrate rule-based symbolic systems and Machine Learning approaches within a framework that takes machinic embodiment seriously (Quinlan 2006; Chalup, Murch, and Quinlan 2007). Finally, I analyze the latest evolution of connectionism in the "deep learning" revolution (Hinton, Osindero, and Teh 2006; Bengio 2009; Bengio, Courville, and Vincent 2013; Arel, Rose, and Karnowski 2010) and examine its utilization in agent-based systems (Mnih et al. 2013, 2015; Nath and Levinson 2014).

At the same time, I implicate the scientific histories of Machine Learning in the processes of how such concepts arose in cognitive science, and, in particular, in the tension that these histories highlight between a representationalist/computationalist and a situated/performative view of the brain (Turing 1950; Searle 1980; Harnad 2007, 2005; Boden 2006, 1996). I do this while contrasting these histories with phenomenological and neurophenomenological critiques (Dreyfus 1979; Maturana and Varela 1980; Varela, Thompson, and Rosch 1991). Cognitive science is a rigorous interdisciplinary field which brings together computer science, philosophy, linguistics, psychology and biology, and thus seems like a logical starting point for examining artificial intelligence while expanding into other fields. In order to understand the relationship between human and machine forms of cognition, I also explore sociotechnical work of researchers in areas such as technocultural studies (Hayles 1999; Johnston 2008; Helmreich 2000), and science and technology studies (STS) (Latour 2005; Pickering 1995, 2010), in order to understand how such flat hierarchizing between human and nonhuman subjects and objects serve to decentre strictly "human exceptionalist" approaches to the links between humans and sociotechnical frameworks for knowledge production.

The third category of writing that I explore in this research comes from media art history and theory as well as agent-based media artworks, examining the manner in which such scientific systems are appropriated by artists. Here, I demonstrate the gaps in art historical accounts surrounding issues of self-organization, adaptation and learning, while attempting to disentangle these terms from their often confused or inaccurate appropriation by art historians and media critics. In doing so, I also pose the central questions of why such techniques are used in art practice in the first place, and what they hope to accomplish.

I begin this endeavor by considering the influence of Cybernetics on art in the 1960s (Ascott 2003a; Burnham 1968; *Software* 1970; Hultén 1968), artificial life art (Penny 2009; Whitelaw 2004; Tenhaaf 2008) and agent-based art (Downie 2005; Mateas 2001; Penny 2000). I pay special attention to perspectives on the figure of the artist-engineer at the crossroads between art and science, specifically as a way to analyze my own practice as an artist trained as an engineeer (Penny 2008; Xenakis 1981a; Xenakis and Messiaen 1994). I compare nonadaptive agent-based installations to ones that use adaptive computational systems, such as Ruairi Glyn's multi-robot installation *Performative Ecologies* (2010), Yves Amu Klein's robotic sculpture *Octofungi* (1996), as well as Stephen Kelly's *Open Ended Ensembles* series (2014—2015). Through this analysis, I try to raise the specific behavioral characteristics of artistic works that use Machine Learning while engaging in larger discussions about the kind of worldview they suggest.

1.5 Methodology

This research follows the framework of "research-creation", a growing set of largely qualitative methodological approaches within the humanities which comes under many variants. Specifically, my methodology aligns with "art-as-research" whereby art practice is embedded in theoretical considerations in a bidirectional network of interactions.

> Art and theory, in effect, are nothing more than two different forms of practice interrelated through a system of interaction and transferences. In this constellation, philosophy neither brings the arts to the point nor does art sensualize philosophical truths; philosophy serves a knowledge-based artistic practice as a point of reference, similar, conversely, to how art might affect theoretical practice. (Busch 2011, 1)

In this context, the process that guided this research is anchored into both material and discursive practices, traveling alternately between thinking and making. Specifically I adopt an iterative design approach inspired from Agile, a methodological framework for software development that largely bottom-up, iterative, and adaptive (Rasmusson 2010). Agile relies on the self-organization of collaborators on a project and values flexibility and adaptability in the planning and decision making process. As an artist who develops open-source software, who is also often working collaboratively on art projects, I have successfully used Agile in the past and have found it a suitable approach for making art that involves software.

Agile rests on the following process:

1. *Break down* a project into small units called *user stories* that describe situations we would like the software to perform. These stories can range from the very abstract (e.g., "I want the software to be able to generate reports") to being very specific (e.g., "I need to be able to export my monthly financial report with the push of a button"). In artmaking, this amounts to establishing artistic intentions and components of the final work.

2. *Estimate* the resources needed to accomplish these tasks (usually measured in days of work).

3. *Prioritize* the list of stories. This is usually where you also establish a production calendar. In an artistic context, where resources are often scarce, I typically choose to begin with components that take less time to accomplish and that give me the most information in regards to questions of technical feasibility and aesthetic effectiveness.

4. *Execute* the plan, updating it as you go. If the project is not moving fast enough, you can either choose to do "less" (which, in media art, could actually be beneficial, as projects using technology often tend to be so complex, that they risk reducing the overall aesthetic effect that was initially aimed for), or you can decide to allocate more resources (hire people or add days to the calendar, if possible).

Agile possesses many interesting features that are relevant to a research-creation strategy. First, it does not presuppose a temporal succession of activities such as: analyze, design, code, test, repeat. To the contrary, it rejects such a method in favor of a continuous model where all of these activities happen at the same time. This is very effective in art-based research, where theoretical and practice questions are never fully separable, and are rather intertwined throughout the whole process. Knowledge is thus constructed as part of an ongoing feedback loop between theory and material practice.

Second, Agile favors an iterative design methodology with short development cycles. In Agile, you usually start by implementing a simple working prototype of the final application, and you add incrementally to it over time. This aspect is particularly well adapted to artistic research, where it is often hard to pin down which aesthetic strategies will work best until one sees a material embodiment of their ideas and intuitions.

Third, Agile values working software as the primary measure of success. Principles such as YAGNI[8] and DRY[9] reduce development costs and prevent overdesign by focusing on making things work. This is a very useful principle in the context of media art, as it is, alas, not uncommon to see artworks in galleries that are plainly defective.[10]

Finally, Agile's planning methodology is adaptive, meaning that it allows for changing plans along the way when faced with reality. Both artists and programmers alike know that it is hard to know how you are going to achieve something, or whether it will really yield the expected results and effects, until you have gotten your hands dirty. This final component is, as expected, intimately linked with many of the theoretical and practical questions approached in this research.

Throughout this research, I rely on records of practice as a way to empirically examine these processes. Through the gathering of documents (notes, diaries, video documentation), as well as introspection and interviews with collaborators, I investigate three (3) artistic projects and the resulting works within which I participated either as solo author or as co-author, thus informing the critical thinking about adaptive systems within an artistic context. The objective is to understand the characteristics and limitations of adaptive agent-based systems by focusing on the reasons which pushed me to use Machine Learning, the ways I have applied these techniques in the creative

[8]YAGNI stands for "You Ain't Gonna Need It" and demands coders not to add functionalities that are not expressely needed.

[9]For "Don't Repeat Yourself", a principle that asks programmers not to repeat information in different parts of their code.

[10]I teach my students that if a new media work functions reliably, even the least of aesthetically compelling media artworks are better than "busted masterpieces".

process, and how the audience reacted to them[11]. I then examine in detail the algorithms that we used, how they were utilized in the research-creation process, and how they affected the outcomes, more specifically in terms of experience.

It is important to highlight the fact that the chosen methodology itself embraces adaptivity as a mode of knowledge generation. The short cycles and the adaptive planning method, allow the various agents engaged in the research-creation process to more tightly adjust to one another, providing a structured yet open-ended frame that gives ground to the emergence of new theories about the world.

1.6 Chapter Breakdown

The following chapters alternate between accounts of practice (chapters 2, 4, 6) and theory (chapters 3 and 5), each chapter responding to the previous one and feeding into the next.

In chapter 2, I discuss my previous work with agent-based systems. Focusing mainly on *Absences* (2008–2011) — a series of environmental interventions using electronic agents set in natural settings — I highlight the research and creation processes that brought me to consider the use of adaptive procedures in agent-based systems, which in turn opened up questions about aesthetics and practice with such systems that lie at the core of this research.

Chapter 3 presents traces the history of adaptive computation and Machine Learning from the 1950s onward. In this first historical overview, I discuss important notions related to agent-based artworks, such as emergence, self-organization, adaptation, evolution, connectionism and Artificial Life. I try to highlight how Machine Learning and adaptive computation operate in these sociotechnical and art historical frameworks, in order to extract the different ways artists have been using them, as well as embracing the challenges that come with such practices. By dissecting the scientific description of learning algorithms and connecting their properties with artistic questions, I establish a comprehensive framework artists and media theorists can use to approach Machine

[11] It is important to mention that at the moment of writing this proposal, the majority of the practice-based aspects of the research have already been carried out. The research will thus mainly consist of reinterpreting the results and bringing them together to create a logical whole.

Learning in works of art.

In chapter 4, I examine *Vessels* (2010—2015), a robotic installation consisting of a swarm of autonomous water vehicles whose collective behavior resembles social interactions in a community of living creatures. The piece involves adaptive mobile robotic agents with complete sensorimotor systems. The significant technological component of the work as well as my direct involvement with the material in this project gives an opportunity to better exemplify the practical considerations of an agent-based artwork involving adaptive methods. The chapter offers a look at how adaptivity plays artistically in agent-based art using the framework developed in chapter 2, exploring the application of Machine Learning as a pathway to generate self-orgainzing, lifelike systems, while contrasting artistic objectives with audience response.

Chapter 5 digs deeper into the notion of behavior from an aesthetic perspective, trying to understand the role and position of adaptation, learning, and emergence in the temporal unfolding of agents' observation-action couplings. I describe how adaptive and evolutionary agent-based systems allow for new morphologies of behavior characterized by the establishment of a second-order relationship to time, one wherein the past affects the future through the transformation of a structure.

Finally, in chapter 6, I discuss the work *N-Polytope: Behaviors in Light and Sound After Iannis Xenakis* (2012), a "spectacular light and sound performance-installation combining cutting edge lighting, lasers, sound, sensing and Machine Learning software inspired by composer Iannis Xenakis's radical 1960s–1970s works named Polytopes", directed by artist and researcher Chris Salter and involving an interdisciplinary team, including myself who created the media behavior modeling and programming. This project allowed me to test a number of Machine Learning and other adaptive algorithms such as Reinforcement Learning and Artificial Neural Networks in a large-scale installation setting involving multiple agents. In revisiting the work of multidisciplinary artist Iannis Xenakis, it provides a good starting point for thinking about the question of the relationship between the spatio-temporal unfolding of adaptive systems and "alloys" of art and science practices (Xenakis and Messiaen 1994).

The last chapter concludes the research by bringing back the questions and providing summaries of the arguments. It then examines the broad implications of the study with respect to the overall areas of study. I end by discussing the limitations of the research and exploring future works that could address some of these limits.

The reader will find additional material concerning some of my own works discussed in this dissertation in the appendices. Appendix A includes references to external web resources such as blogs and video documentation, while appendix B contains full-page images of these works.[12]

[12] Many of these images are repetitive of figures found in the core of the dissertation, in larger format.

Chapter 2

Absences

> The best bridge is one that just stands there, whatever the weather. Cybernetic devices, in contrast, explicitly aimed to be sensitive and responsive to changes in the world around them, and this endowed them with a disconcerting, quasi-magical, disturbingly lifelike quality.
>
> – Andrew Pickering, *The Cybernetic Brain*

I need to wait for it everyday. First thing in the morning, I put the circuit out under the sun, letting it wake up, giving the batteries a chance to recharge. The days have a strange shape. I seem to be waiting for the sunset to come for the whole day, measuring the voltage increase as the day goes, doing some internet [sic], writing some code, running simulations.

Then, when it comes, I'm always late. Around 4 o'clock, I got to [sic] restart the circuit with the new, enhanced program. Even though I'm waiting for it all day long, I'm always running after it, coming back home with the scooter, getting the program to compile, checking if everything is fine. Then I wait, carefully looking at the evolution of the little indicator LED.

It goes very fast. In about an hour, the sun dives down to the horizon, setting the sky on fire. You can feel it in your skin.

Then the chill comes.

> I'm not used to this. Dealing with reality. I can debug only once a day. For a programmer, this is nightmare. But I like it, that pace, that rythm [sic]. Ageless and uncontrollable, it is incredibly comforting.[1]

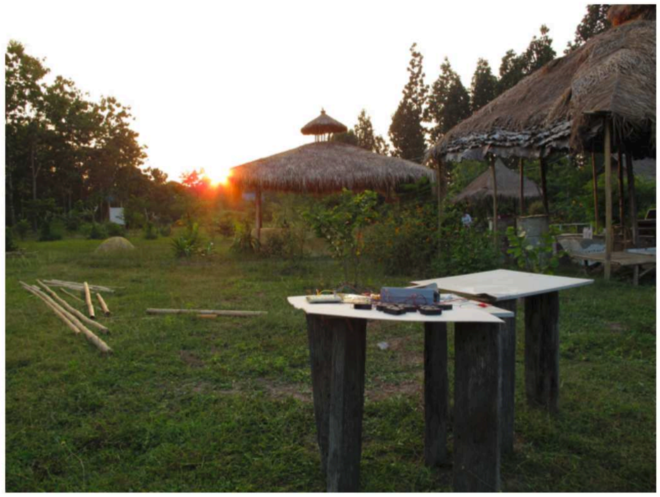


Figure 5: Waiting for the sunset at the ComPeung art center (Doi Saket, Thailand.

I posted this entry on an online diary in December 2008, five years after I had finished my computer science degree. What was I doing there, an engineer trained in Artificial Intelligence, now in a rural area of Northern Thailand, struggling with the elements? What led me so far away from home and my zone of comfort?

As explained earlier, my first few years as an artist had led me to create interactive works that were meant to represent different kinds of social or natural phenomena. However, after completing my Communication degree in 2007, I begun to question that approach, and started exploring ways to act directly in real life through small electronics agents. During a residency in the Netherlands, I created *Drift*, an interactive sound device that attempted to adapt to its surroundings (fig. 6). In the Fall of 2007, I started two projects in parallel that would allow me to further experiment with these ideas.

The first project, *Accrochages* (2008), was co-authored with Samuel St-Aubin, a Montreal-based artist whose work aims to rethink everyday life objects through the creation of autonomous devices.

[1]Excerpt from the Absences project blog (http://absences.sofianaudry.com/en/node/33).

Samuel St-Aubin is a self-taught artist who excels in hardware design, physical computing and rapid prototyping. His extensive knowledge of electronics, combined with my programming skills, would give us an opportunity for sharing approaches and learning. Inspired by psychogeographic techniques of "détournement" ("drifting") and street art, our work consisted of imagining and designing low-cost electronics devices that could be distributed in the urban space, using "simple means to give new qualities to the city environment by creating different interactive situations".[2]

Figure 6: *Accrochages* (2008). Art souterrain, Montréal, Canada. Photo by Alexis Bellavance.

The second project, called *Absences* (2008—2011), involved a series of five (5) electronic interventions where artificial-life agents were installed in outdoor environments. Taking shape at the frontier of new media and environmental art, it proposed a meditation on solitude and association, interaction and adaptation, natural and artificial, biological and inanimate. Each intervention consisted of the creation and installation of autonomous electronic devices in various ecosystems. These artificial agents acted and reacted within their specific environment.

In *Absences*, I set out to meddle directly with natural processes. The very concept of interactivity, largely explored in my past work, was put under scrutiny: these systems were no longer meant to interact with human spectators, but with a whole ecosystem of nonhuman agencies. One of my core interests with the series was to subvert the accepted notion of technology as something

[2]Source: http://accrochages.drone.ws/

"useful" that embodies man's unbounded control over nature. Instead, here, I sought to place frail technological systems in the outdoors, with the modest goal of giving them the ability to survive within their habitat — an objective which would prove to be much harder than I originally thought.

Though most of my previous works were interactive, one of the important choices that I made was to purposefully keep humans out of the equation. The first reason was conceptual: interacting specifically with humans would have countered the spirit of the project itself, which called specifically for a decentering of the human subject in our view of how technology and nature interoperate. Having worked on many interactive projects in the past, I wanted with *Absences* to subvert the human-centric concept of interaction by "interacting with nature".

The second reason was more practical. Humans are extremely complex agents and the way they choose to interact with a piece is equally unpredictable. As I was already taking a risk by intervening in the outdoors — a rather hostile place for electronic entities — it felt like adding the extra challenge of accounting for human behavior in the design was a bit excessive and would limit my freedom. From my perspective, natural phenomenon possessed a more "predictable" dynamics, which would facilitate the integration of the devices in outside milieus. This an assumption would turn out to be quite overblown.

I created *Absences* as a research-creation program which would allow me to move beyond a practice based on representations, and instead towards one anchored in interventions and performances. The projects of the series inspire important questions about technology, nature, and nonhuman agency. Can artificial agents "survive" in nature (and how)? What is the aesthetic effect of these agents? How are they connected to art, science, beauty and truth? How do they refine and redefine notions of agency and behavior in both science and art?

Absences marks a turning point in my research-creation practice as it resurrected my interest in adaptive systems, fostering the fundamental questions at the core of this dissertation. One of the important aspects of *Absences* — and the reason why I chose to specifically dedicate a chapter to the project — is that there is a clear progression in the kinds of behaviors that were produced in the project, from rule-based nonadaptive systems to self-regulated, self-organizing, and adaptive processes. In particular, the last two interventions were adaptive, the fourth device being driven by

(a) *First Absence* (2008).

(b) *Second Absence* (2009).

(c) *Third Absence* (2010).

(d) *Fourth Absence* (2009).

(e) *Fifth Absence* (2011).

Figure 7: The five interventions of the *Absences* series (2008–2011).

a very simple feedback system, while the fifth and last agent was governed by a Machine Learning algorithm using neural computation.

In this chapter, I describe my creative process in the design of the *Absences* series, which re-activated my interest in Adaptive Computation and Machine Learning. I first recount my failed attempt at building an agent activated by sunset during my first intervention in Northern Thailand, showing how it gave rise to the need for integrating adaptive behaviors in my work. I then report my use of different kinds of adaptive and learning algorithms, with increased complexity, in three of the other interventions of the series. Finally, I discuss how *Absences* opened up the broader set of questions that are addressed in this research.

2.1 The Need for Adaptation

One of the decisive moments in *Absences* happened while designing the first intervention, a device that would only activate at sunset by inflating artificial "fruits" using small air pumps. Whereas the device had solar panels and light sensors which allowed it to perceive incoming light, I wanted to prevent it from being "fooled" by passing clouds or dust that could potentially accumulate on its sensors. Thus, I gave the agent some extra sensors in the form of thermoresistors. The algorithm, which I programmed "by hand" over the course of several weeks, looked not only at absolute values, but rather at the variation in both light and temperature through the day, on the assumption that the general slope of change in both temperature and light are more robust measurements (in particular with respect to seasonal variations).

Figure 8: *First Absence* (2008), ComPeung, Doi Saket, Thailand. Photo by Sofian Audry.

The excerpt from the project's blog that began this chapter describes the creative process undergone during that period. It shows how I myself became engaged in an adaptive procedure, making adjustments from sunset to sunset, my agency intertwined with nature's immutable cycle.

I installed the module around December 2008, on a small tree located on the ComPeung residency center's land.[3] For a few days, I made sure the piece worked as I had so carefully designed it

[3]For reference, read: http://absences.sofianaudry.com/fr/node/38.

to, marking every sunset with its daily behavior. In January and February, I became more occupied with my second intervention and did not look too closely at it.

At the end of my residency in late February 2009, the center held a public event to showcase the work. As everyone anxiously waited for the device to start, the sun slowly descended to the horizon. After about thirty minutes, the sun was almost gone with no reaction from the "artificial tree", forcing me to manually start it by flashing the microchip with an emergency program.

In the space of a few months, the initial conditions had changed and the adjustments I had so carefully made in December to the light and temperature threshold, were no longer appropriate. The algorithm, constrained by these hardcoded parameters, was utterly unable to *adapt* to seasonal variations in daylight. That agent was therefore incapable of "surviving" in its environment, in the sense that its behavior was unable to maintain itself through environmental changes. It was unable to be "alive" — or at least stay alive — in these conditions without adaptation.

More importantly, the device's aesthetic identity as a whole required adaptation in order to exist temporally. While such an installation could have easily worked for an extended period of time in a controlled setting such as a gallery space, the fixity of its design was hereby brutally exposed through the rich variability of nature's complexity.

I claim that learning and adaptive systems suggest a complete change in paradigm in regards to technology and how it operates in the world. Technologies of the past and present are immensely nonadaptive: they are driven by a human-centric ethics that seeks to control nature (Pickering 2010). Quite paradoxically, current-day Machine Learning has not really escaped that paradigm, being used for the most part for pattern recognition purposes in attempts to efficiently solve concrete, measurable "problems"; to gain more control over outcomes.

2.2 A Narcissistic Self-Regulating System

In the meantime, I had already finished creating the *Second Absence* (2009), a small device consisting of a simple input-output system involving an LED and a photoresistor enclosed in a glass bottle. The device, which was installed deep in the Thai rainforest and only activated during nighttime, was driven by a minimalistic self-regulating mechanism where it tried to adjust its light level (actionable through the LED) to counter-balance its perception of light (through the photocell), as if rapidly reacting to its own reflection.

The algorithm went as follows:

1. Let x be the value of light as read by the photoresistor, normalized and remapped to $[0, 1]$.
2. Set the LED value to $(1 - x)$ (i.e., the opposite of perceived light).
3. Wait for some time.
4. Go to step (1).

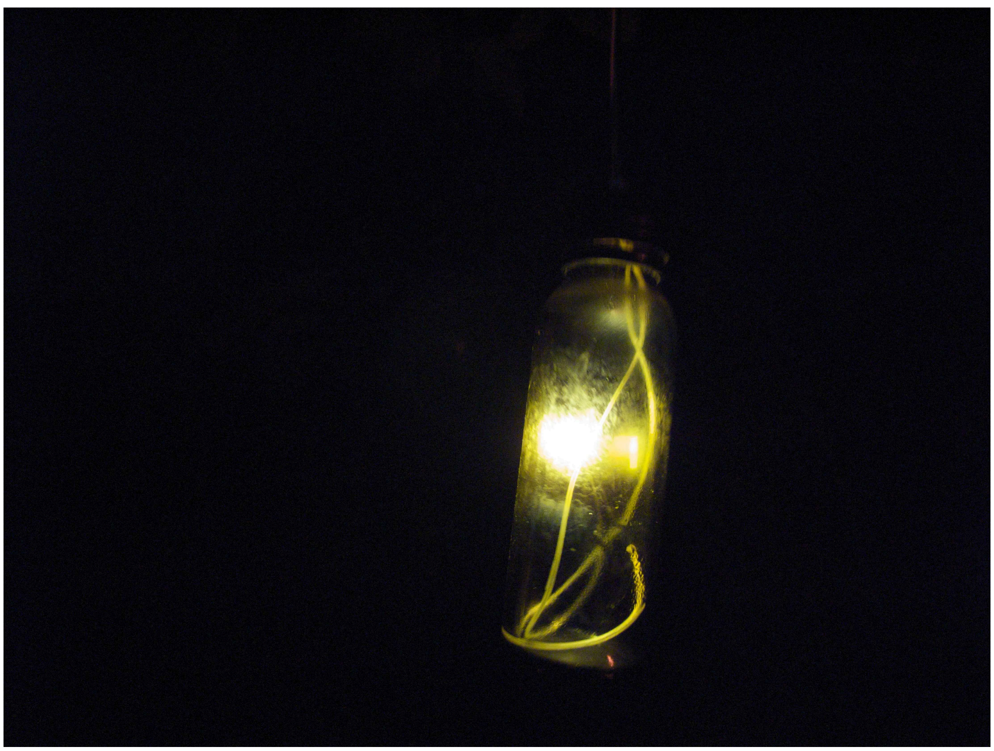


Figure 9: *Second Absence* (2009), Mae Kuang reservoir, Thailand. Photo by Sofian Audry.

The process results in a rapidly, yet unstable flickering light, as the agent iteratively adjusts its actions to its perceptions. It stabilizes after a few seconds, asymptotically reaching a state of equilibrium. When the agent notices that its perceptions match its actions, it quickly gets "bored" and moves into a "sleeping" state, represented by a simple sinusoidal oscillation.

This second intervention displayed a rather formal process that at the same time contained a simple kind of adaptation called "self-regulation" through a closed feedback loop. We will see in chapter 3 how self-regulation and negative feedback systems are the first principles behind adaptive systems as they were defined by first-order Cybernetics.

2.3 Surviving in the Wild

The *Fourth Absence* (2009) used a similarly simple self-regulating process, however this time running in an *open* feedback loop, fully engaged with the natural elements.[4] The project directly engaged with the situation of allowing an artificial agent to "survive" in a hostile setting, exploiting the aesthetic potential of its own energy management. The agent consisted of a three-meter high tube, most of which was buried underground to protect the circuit and batteries from the extremely low temperatures of Winter. At the top, two solar panels allowed the device to recharge its batteries, with an efficiency that was expected to vary highly through the seasons.

> Energy management is a concrete example of acting within nature and a recurring issue in the project. I will here focus on a kind of device that have insufficient access to resources and thus needs to alternate between periods of activity and dormancy, such as is the case for most real-life organisms. How can such a device reach its specific goals in balance with the available energy resources?
>
> A solution to that problem was developed during my stay near the Arctic (Yukon, 2009). I built a device that produced a sound at a specific pace. Between each sound emission, it would switch to a sleep mode, consuming almost zero power. The massive changes in day length in the region throughout the year requires it to adapt its frequency accordingly. The right frequency cannot be computed analytically since it depends on many unkown [sic] factors (such as the temperature and the precision of the sensors).
>
> I addressed this issue by relying on a very simple adaptive algorithm that updates the frequency of appearance of the action (in this case, emitting the sound) based on the

[4] For the sake of focusing on the most relevant works of the series, I will not discuss the_ Third Absence_ in detail here. Suffice it to say that it played with the question of feedback and self-organization with multiple agents using a very simple formal algorithm, using nature's own indeterminacy as its main strategy. For more information, please consult the following video documentation: https://vimeo.com/46469372.

measured batteries [sic] power (voltage). If too much power is available, the frequency is slightly increased, rising the energy consumption. If there is not enough, it is reduced in a similar fashion. (Audry 2010, 2–3)

Figure 10: *Fourth Absence* (2009), Dawson City, Yukon, Canada. Photo by Sofian Audry.

The intervention thus applied a principle very important to early Cybernetics: homeostasis, the mechanism by which organisms are able to keep some of their inner variables stable through time by acting in purposeful ways with respect to their inputs. In this case, the charge of the battery is the variable that is kept stable by adjusting the period at which the sound is played.

As in the *Second Absence*, the behavior is based on a simple feedback procedure — which as we will see in section 3.1.1 is called *negative* feedback (Wiener 1961) — that updates a single parameter w that controls the period T (expressed in hours) between each action (i.e., sound emission) using the following formula:

$$T = \frac{1}{1 + \exp(-w)}$$

5

When night comes, the agent compares the voltage that remains in its batteries (V_{batt}) with a target voltage (V_{target}), producing the following error value:[6]

$$E = \frac{V_{batt} - V_{target}}{V_{max}}$$

This error becomes a way to assess the success of its strategy in choosing parameter w. The agent then makes a step-wise adjustment to w, adjusting it to try and lower the error on the next day, using a simple learning rule:

$$w \leftarrow w - \eta[(1 - T)T]E$$

This expression corresponds to a very common form of adaptive procedure called a *stochastic gradient descent* (Bishop 1995). Parameter η is a small positive value called a learning rate which controls the speed of adaptation. It needs to be set by hand to a value large enough to ensure change, but small enough to smooth out natural variations in the data.[7]

Without entering into details, consider the case where too little energy has been consumed over the day, with $V_{batt} > V_{target}$, hence yielding a negative error ($E > 0$). Assuming the periodicity of the actions is at least partly responsible for the situation, one can see how w will be *decreased* because $\eta[(1 - T)T]E > 0$. If that is the case, then the period T will also be decreased, which is

[5] This formula corresponds to the *sigmoid function*, an activation filter commonly used in Artificial Neural Networks. One of its properties is that it maps values between 0 and 1 in a non-linear fashion (hence the expression ensures that T stays between 0 and 1 hour). See Bishop (1995) for more information.

[6] The difference is divided by the maximum possible voltage of the batteries (V_{max}) to ensure it stays between 0 and 1.

[7] Though this parameter is set by hand, it is relatively robust, meaning that a wide range of values for it will still allow learning to occur. There is of course a tradeoff between speed of convergence and precision.

For example, if we consider the case of the Fourth Absence, choosing a high learning rate will result in the agent reacting very abruptly to day-to-day variations (which might be due to natural noise), whereas choosing a low rate could prevent the agent from adapting fast enough to seasonal changes.

exactly what we want in this case: we would have had enough energy to produce the sound more often, hence we want to increase the pace by lowering the time between each action.

2.4 Adapting to Conflicting Desires

The issue of adaptation proved to become an increasingly important aspect of *Absences*, both practically, conceptually and aesthetically. For the final intervention, I decided to extend the idea of an adaptive mechanism by trying to integrate a kind of Machine Learning procedure called Reinforcement Learning (Sutton and Barto 1998) (see section 3.2.1 for more detail) by designing a sensorimotor system animated by its own set of conflicting desires, which would be able to adapt and learn from its actions.

Figure 11: *Fifth Absence* (2011), Catalonian Pyrenees, Farrera, Spain. Photo by Sofian Audry.

For this project, I imagined a robot that could control the orientation of one or more solar panels which would make up its body using servomotors. These solar panels would be used both as sources of energy and as photosensors. The robotic agent would start with a “blank mind”, not even knowing the relationship between its movements and the orientation of its solar panels. It would simply explore its environment and, little by little, through trial and error, establish a correlation between its actions and perception.

In December 2009, I had successfully used a Reinforcement Learning algorithm in the realization of a very simple adaptive agent. Reinforcement Learning is an approach in AI that allows an artificial agent to learn an optimal behavior from a series of actions based on observations, through an iterative process of trial and error. In each action, the agent receives a reward or punishment in the form of a number — positive or negative — which allows it to guide its future choices.

I found this approach exciting because it seemed to be an efficient way to design self-organizing, emergent, potentially surprising behavioral patterns while giving some “guidelines” for the agent to follow. The resulting behavior is not determined in advance: it is chosen by the agent itself, according to its particular context and the rewards it receives, through its interactions with the environment. The practitioner can thus work with the context (i.e., the inputs and outputs afforded by the system) while encoding the desires of the agent through rewards and punishments. Yet, as the agent itself determines its best strategy to maximize its rewards, the learning process holds the potential to generate unexpected, possibly surprising behaviors.

The agent was anchored to a cliff in the Catalonian Pyrenees, floating above the void. I encoded a very simple reward function which rewarded the agent for looking away from the sun while heavily penalizing it for running out of batteries. In so doing, I I put the robot in a tension between two conflicting choices, forcing it to navigate along the thin ridge between need and desire, slowly adjusting to find its own balance in the world.

2.5 Towards Adaptive Systems

Absences was a first attempt to directly intervene in uncontrolled, so-called “natural” ecosystems through performative, embodied artificial entities. The project, which was articulated as a three years research-creation program, follows the artistic tradition of environmental art, which is itself a particular form of conceptual art.

As I struggled to integrate such embodied artificial agents in changing, often unstable environments, I was pressured and inspired into using adaptive systems, as is summarized in table 2. I began exploring these concepts more profoundly over the next several years with projects such as *N-Polytope* (2012), *Plasmosis* (2013), *Archipelago* (2014), and *Vessels* (2015). Through these works, I reconnected with the aesthetic dimension of Machine Learning that lead me into the field in the first place; only in this artistic context, instead of seeing an error rate get lower, there were real-time, media effects that unfolded in time and space.

Table 2: Evolution of systems in *Absences*.

Title	Year	Description of system
First Absence	2008	Nonadaptive (but should have been).
Second Absence	2009	Nonadaptive but included a process of self-regulation through a feedback loop.
Third Absence	2010[8]	Nonadaptive but involved a multi-agent communication loop that could be seen as a self-regulated system.
Fourth Absence	2009	Adaptive through a simple feedback system allowing the piece to "hibernate" (i.e., adjust its consumption to available resources).
Fifth Absence	2011	Adaptive through a Reinforcement Learning system that made use of an Artificial Neural Network.

This being said, *Absences* opens up a can of worms, provoking larger questions around the use of such techniques within artistic practice. How can the different components and properties of Machine Learning algorithms be exploited by artists? How do they affect the aesthetics of a piece? How is adaptation related to autonomy, emergence, self-organization and self-regulation? How are these concepts connected both aesthetically and historically? How do agent-based systems work aesthetically? What effects do they have on audiences? How can/does adaptation influence these effects? What distinguishes behaviors produced by adaptive systems from those produced by self-organizing and rule-based systems?

To address these important considerations, we must first understand the sociohistorical framework surrounding adaptive and Machine Learning methods. The techniques I have employed in

Absences were taken largely out of context, as I simply applied my own knowledge derived from computer science. In the next chapter, I strive to understand where these approaches come from, to extract them from their pure technical meaning and bring them into a broader interdisciplinary domain, I do so by tracing agent-based and adaptive systems back through the history of computer science from the 1950s onward while simultaneously establishing parallels with media art history.

Chapter 3

Towards a Practice of Machine Learning in Agent-based Art

> I suspect that the "aesthetics of intelligent systems" could be considered a dialogue where two systems gather and exchange information so as *to change constantly the states of each other*.
>
> – Jack Burnham, *The Aesthetics of Intelligent Systems*

> The great difference between magic and the scientific imitation of life is that where the former is content to copy external appearance, the latter is concerned more with performance and behavior.
>
> – Grey W. Walter, *An Imitation of Life*

This chapter aims to provide a strong contextual ground for this thesis. It introduces the fundamental concepts that are studied in the research, such as systems, agents, behaviors, self-organization, emergence, adaptation and Machine Learning. Tracing through the history of adaptive and learning systems since WWII in both science and art, I highlight how these concepts travel between scientific and art historical frameworks, trying in particular to articulate how artists utilize these concepts and the challenges that come with such practices. Finally, I discuss the main components that constitute Machine Learning algorithms, exploring ways they have been, or could be used for artistic expression.

The first part of the chapter explores the history of adaptive and learning systems following a more or less chronological path. One should however be aware of the intricate ramifications that run synchronously to this story. I start by exploring so-called "first-order" Cybernetics (1946–1950s) which is associated with the appearance of early connectionist models (1950s). I then describe the appearance of the field of Artificial Intelligence in the 1950s which came in opposition to Cybernetics and connectionism (1956–1974). The emergence of Cybernetics and AI in the 1950s is associated, in the 1960s, with the rise of new art forms which art historian Jack Burnham has described as "systems aesthetics" (Burnham 1968), and whose larger genealogy lies in artistic movements such as conceptual art, cybernetic art, information art, algorithmic art, etc. I then examine the revival of connectionism in the 1980s which I associate with the emergence of Machine Learning and its development and popularization in the 1980s–1990s.

The second part of the chapter examines the intrinsic properties of Machine Learning systems in an effort to delineate unique artistic strategies that can be used to exploit them. I summarize the different dimensions that define Machine Learning algorithms in the scientific literature, such as the tasks they can solve (supervised, unsupervised and reinforcement), the model being used (neural networks, genetic algorithms), the evaluation criterion (measuring the performance of the system) and the learning process itself. I discuss each of these properties from both and aesthetic and practical standpoints, exploring how such techniques are utilized in artistic works, and based on this knowledge suggest new possibilities.

3.1 Historical Context

History is imbued with a fascination for the possibility of humans to artificially fabricate life. Many stories from Antiquity display artificial, humanoid creatures: Ovid's Pygmalion, who fell in love with a statue of his own making, brought to life by Venus, is perhaps the most well-known of them (Ovid 2008). The figure of the Golem, a humanoid creature made of clay, brought to life by the name of God, appears in early talmudic mythology (Idel 1990).

There exist several records of mechanical automata in the ancient world. One of the first documented example is a steam-activated "pigeon" constructed by mathematician Archytas of Tarentum (circa 400–350 BC). In China, a mechanical orchestra was allegedly built for the emperor during the Han dynasty around the 3rd century BC. In the 13th century, Muslim inventor al-Jazari created a series of moving peacocks for the royalty of the Urtugid dynasty in Mesopotamia. French inventor Jacques Vaucanson, who created the first completely automated loom, designed many life-imitating automata. His most famous work, *The Digesting Duck* (1739), was an artificial bird made of more than 400 parts who could move, drink, eat, and defecate.

These examples are only a fraction of the numerous life-imitating machines designed in both the Eastern and the Western world from the Antiquity to the mid-XXth century. A common characteristic of these mechanical devices is that they were always dedicated to a single set of tasks and could not be easily modified and/or re-purposed to accomplish another one.

But a change in paradigm takes place in the period following WWII. Increased interest in military applications of computation such as ballistics and cryptography led to the appearance of the first general-purpose computers in the late 1940s. Contrary to mechanical automata, which were usually able to address very specific problems such as manufacturing textiles or performing simple arithmetics operations, computers are *programmable*, which means that they are theoretically able to perform almost any kind of algorithmic symbol manipulation.[1] It makes them uniquely powerful, which led many at the time to think that computing is universal and could theoretically model any kind of process found in nature, including animal behavior and human cognition.

Several accounts of the history of Artificial Intelligence and Machine Learning exist from both humanities and social sciences (Hayles 1999; Whitelaw 2004; Johnston 2008; Penny 2008; Clarke and Hansen 2009; Pickering 2010; Halpern 2014; Shanken 2015) as well as computer science (Brooks

[1]There are theoretical limitations to the power of computers, as was first revealed by Kurt Gödel in his famous incompleteness theorems, published in 1931 (Van Heijenoort 1977). Five years later, inspired by Gödel's work, mathematicians Alonzo Church and Alan M. Turing almost simultaneously demonstrated the impossibility of writing a generic algorithm for solving the "halting problem" – or Entscheidungsproblem – which is the problem of automatically reading another computer program and deciding whether it will stop (halt) or run undefinitely (Church 1936; Turing 1936).

1999; Sutton and Barto 1998; Duda, Hart, and Stork 2001; LeCun, Bengio, and Hinton 2015; Medler 1998; Nilsson 2010). Whereas the latter accounts from computer science have the advantage of offering an insider's perspective over the history of the field, it traditionally focuses on the evolution of techniques, often neglecting to contextualize the sociocultural dimensions. As for the former in the humanities and social sciences, they seem to suffer from the inverse illness, as they bring larger considerations into the picture but often fail to understand the science itself and are thus prone to misrepresenting it.

As an interdisciplinary scholar trained in Humanities, Media Art and Machine Learning, I aim to bring some clarity into the debate by developing my own story of adaptive systems since the post-war era. In particular, I want to focus on Machine Learning systems, analyzing them from the perspective of the history of science as well as art history, articulating origins and developments of cultural imaginings surrounding artificial adaptation and the role it plays in contemporary art.

One of the most important challenges lies in the difficulty to trace techniques used by artists, as works using adaptive systems are scarce and often poorly documented. For example, many artists use extremely general terms when describing the techniques employed in their works, such as "neural networks", "ecosystems", or "evolutionary systems".[2] An important aspect of my contribution here is thus to disentangle which methods were actually used in order to connect these works with their scientific practices.

3.1.1 Cybernetics

It is largely accepted that contemporary concepts about artificial agency and adaptive systems such as AI, Machine Learning, and Neural Computation, originated in the early 1940s with the Macy Conferences on Cybernetics (1946–1953). A set of ten interdisciplinary gathering chaired by neurophysiologist Warren McCulloch, these conferences brought together mathematicians, psychiatrists, psychologists, biologists, social scientists and computer engineers, with the ambitious goal

[2]For example, compare the use of the term "neural network" in the work of Yves Amu Klein (Klein 2014) — which refers, in fact, to a specific kind of unsupervised neural net called self-organizing maps — with the dome performance *Bio-inspire* by Turkish audiovisual artists Bahadır Dağdelen and Yusuf Emre Kucur who provide a very elusive description of the kind of networks they are using, with no scientific reference that would allow one to understand the kind of technique they are putting into action (Dağdelen and Kucur 2016).

of constructing a general theory of the human mind (Dupuy 2000).

The conferences revolved around the organism and its relation with its environment. At the first gathering in 1946, Warren McCulloch, following on his recent research with logician Walter Pitts, had shown how propositional logic could be modeled by simple artificial neural networks. In their 1943 paper on neural networks, the authors had proposed a simplified model of neuron activity where each brain cell is in either one of two states at any given time (on/off, true/false, 0/1) (McCulloch and Pitts 1943). The neurons are connected using synapses which are either excitatory or inhibitory.[3] The alleged "all-or-none" neural activity, thus reduced to on/off mathematical components, allows "neural events and the relations among them" to be treated using logical calculus (1).

In January 1943, Arturo Rosenblueth, Norbert Wiener and Julian Bigelow published "Behavior, Purpose and Teleology" where they presented a teleological model of human and animal behavior which would also be shown at the first Macy Conference in March of the same year (Rosenblueth, Wiener, and Bigelow 1943). They defined behavior as a transformation in the organism related to its environment (1). Recognizing the extreme broadness of the definition, they developed a hierarchical taxonomy of behavior, classifying animal behavior as active (as opposed to passive), purposeful (as opposed to random) and teleological.

Teleology is key to understanding the origins of contemporary notions of adaptation. It is tightly connected to the notion of feedback, a concept that would become a central component of Cybernetics (Wiener 1961). The term *feedback* comes from engineering where it has two meanings. The first sense of the word, called *positive feedback*, refers to a property of a system where "some of the output energy of an apparatus or machine is returned as input", such as in an amplifier circuit.[4] Thus, when talking about (teleological) feedback, the authors rather refer to the second sense of

[3]In neurological terms, an excitatory synapse increases the likelihood that the post-synaptic neuron will fire when the pre-synaptic neuron does, while an inhibitory synapse decreases that likelihood. Inhibitory synaptic connections play an important regulatory function in the brain.

As an example, epilepsy seizures are the result of a dysfunction in the inhibitory mechanisms in the brain which causes neurons to fire erratically due to unregulated excitation.

In McCulloch & Pitt's design, inhibitory inputs are absolute, meaning that if a neuron receives many inputs, inhibition will take precedence over excitation, thus preventing the post-synaptic neuron to fire.

[4]Economic collapses are usually caused by such positive feedback loops: as people lose their trust in the market they begin selling their assets, which causes more people to lose their trust in the market, and so on.

the word, called *negative feedback*. Here, the difference between the goal and the outputs (i.e., the error) is sent back to the inputs, allowing the system to correct its course. Thus, for the authors, teleological behavior refers to the changes undergone by a system that tries to reach a given goal in its environment, constantly adjusting using a feedback loop.

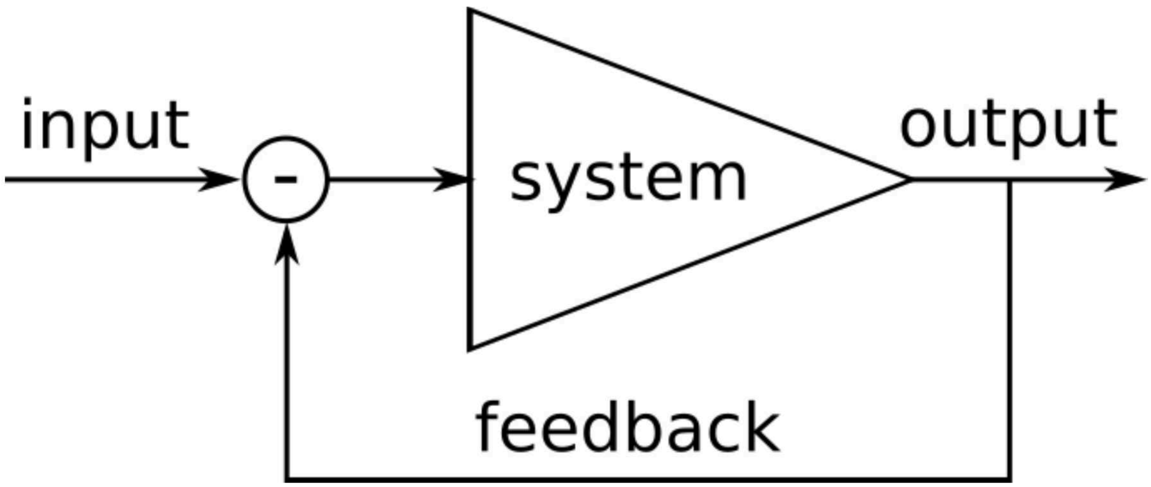


Figure 12: Schematics of Norbert Wiener's negative feedback loop. Some of the output energy of the system is subtracted from its input as a way to adjust its course of action.

For example, the *Second Absence* project described in the previous chapter (section 2.2) contains a simple negative feedback mechanism whereby an agent's actions (changing the intensity of a diode) directly influence its own sensory perceptions (a photoresistor) and, in turn, trigger an inverse reaction as the system tries to compensate for either a lack or excess of luminosity.[5] Interestingly, the natural behavior of the system is to oscillate with a large amplitude at first, but to slowly and steadily move towards stability. As we will see, most Machine Learning training algorithms use forms of negative feedback as a way to search the space of possible solutions to a problem.

Renowned Duke University literary critic N. Katherine Hayles highlights another essential element of these early years of Cybernetics in Claude E. Shannon's theory of information. Shannon, a mathematician and engineer, had worked as a cryptographer during the Second World War. In his seminal 1948 paper, "A Mathematical Theory of Communication", Shannon establishes a

[5]In contrast, the *Third Absence*, where the sonic perception of a single agent triggers a chain reaction as it sends the sound to the next agent, perhaps exemplifies a positive form of feedback.

clear separation between the message and the signal that encodes it. He states that the *meaning* that the message conveys is unimportant to the engineering problem of communication, which is rather concerned about its *probability* of appearance. Shannon thus defines information as an immaterial measurement of uncertainty that has nothing to do with significance; a "pattern, not a presence" (Hayles 1999, 33).

Rosenblueth and Wiener's definition of behavior and Shannon's theory of information would both have a major impact on Cybernetics. They are related to another foundational concept in the field: *homeostasis*, a property of a system that constantly adjusts the output of an organism such that it maintains a state of stability using an adaptive criterion embedded in a negative feedback loop.[6] A thermostat is the perfect example of a simple homeostatic system. It tries to regulate the ambient temperature using negative feedback, switching the furnace on when the temperature is too low, and off when it gets too high.

The perfect embodiment of homeostasis can be found in Ross Ashby's *homeostat* (Ashby 1957), a physical device that can adapt to its environment using a feedback loop. The homeostat was presented by Ashby at the ninth Macy Conference. It is an electrical device made of two parts. The first part consists of four units. Each unit has an electrical magnet that can deviate from its central position. The deviation of each magnet is converted into an electric current which is sent as an input to the other three units. Within the system, all units are interconnected. Moreover, each unit sends its electric output as a feedback input to itself. The inputs control the activity of coil relays that move the magnet in such a way that the deviation of the magnet is roughly proportional to the sum of the currents (Ashby 1954, 95).

With appropriate tuning, the device displays extreme stability. "If the field is stable", Ashby explains, "the four magnets move to the central position, where they actively resist any attempt to displace them. If displaced, a co-ordinated activity brings them back to the centre." (96) For Ashby, this "ultrastability" found in homeostatic systems is a necessary condition of life (110).

The *Fourth Absence*, described in the previous chapter, is an example of a simple homeostatic device. In effect, it tries to maintain a stable variable (its battery level) over which it has only

[6] The concept was first described by physiologist Claude Bernard in 1865 (Bernard 1957) and the term *homeostasis* was coined in 1928 by Walter Cannon (Cannon 1928).

indirect control by tentatively adjusting its actions (the frequency at which it plays a sound) so as to adapt to seasonal variations (cf., section 2.3).

3.1.2 Early Connectionist Models

Early after the war, the science of Cybernetics in the Macy era thus started designing rudimentary adaptive, self-regulated systems able to "stay on course", moving towards a definite target by making micro-adjustments to their internal structure. Of particular interest is the self-organizing nature of Cybernetics devices such as the homeostat, which also suggests a perspective on human cognition. From this viewpoint, memory functions not as a definite trace image (like a "snapshot") that can be retrieved through some kind of addressing mechanism, but rather as the real-time, dynamic relationship ongoing between a distributed set of control units.

In 1949, Canadian psychologist Donald O. Hebb proposed a revolutionary model for human neural networks that went along similar lines. He claimed that as brain cells subject to certain types of stimuli fire together, they also increase their likelihood of firing together in the future when subjected to similar stimuli, thus forming self-organized assemblies of neurons (Hebb 1949). This principle, a "form of connectionism" (xix) which would come to be known as *Hebbian learning*, views human memory as a subsymbolic, distributed, self-reinforcing process, rather than as a collection of coded representations that would somehow be stored in the brain.[7]

Building upon both Hebb's findings and cybernetician models of the brain such as Ashby's homeostat and McCulloch and Pitts' logical neural nets, Frank Rosenblatt proposed in 1957 one of the first adaptive connectionist devices: the *Perceptron* (Rosenblatt 1957). The perceptron is a simplified model of a human neural network in the shape of a thresholded linear function[8] that is able to classify a pattern in one of two categories.

[7]Physiological theories of learning similar to Hebbian learning had been around since the 19th century. For an in-depth historical review, see Cooper (2005).

[8]A linear function is a polynomial function of degree zero or one. In the one-dimension case, such a function has the form:

$$y = mx + b$$

which once drawn on a cartesian plane shows as a line with slope m, hence the attribute "linear".

The Perceptron uses such a linear function, but in the more general n-dimensional domain, where n is the number of inputs:

It maps a set of typically binary data (input neurons) to a binary output (output neuron) using a layer of parametric values called weights (representing the synapses). The weights are usually initialized randomly. A simple training procedure allows the perceptron to adjust its weights based on a series of example inputs for which the expected output is known.[9]

For example, suppose that we want to differentiate handwritten letters that are either "A" or "B" using a Perceptron. We create a database of multiple 8x8 black and white images of handwritten A's and B's. Each such image can thus be represented as a vector x of 256 dimensions $x_1, \cdots, x_{256}$, each being assigned a value of -1 to represent black pixels and a value of 1 for white pixels. The model possesses a weight w_i that is associated with each of the 256 inputs of an image, which is usually initialized randomly.[10]

To compute the class of a given input as predicted by the Perceptron, we feed it one of the examples by copying the values of one of the images to the network's inputs x_i, multiplying each of these values by its corresponding weight w_i and adding the results:

$$o(x) = \sum_{i=1}^{256} w_i x_i + b$$

Thresholding the value at zero (0), we classify this image in the "A" category if the resulting sum is negative, and into the "B" category if it is positive. Let y be the category predicted by the network:

$$o(x) = \sum_{i=1}^{n} w_i x_i + b$$

where w_i are the weights (synapses) associated with inputs x_i (eg., each pixel in a black and white image), while b is a "bias" weight. In that case, the function can be represented geometrically as a hyperplane in n dimensions that splits the space in two distinct regions representing the two classes that we try to distinguish.

The result (a scalar) is then "thresholded" to yield a binary output (representing the true/false category the perceptron is trying to infer from the input data): the category will be 1 if the output $o(x)$ is positive, and 0 if it is negative.

[9] It is worth mentioning that the perceptron was invented around the same time as another connectionist network inspired by the McCulloch-Pitts model, the Adaptive Linear Element or ADALINE (Widrow and Hoff 1960), which uses a similar learning rule.

[10] There is also an additional bias weight b that needs to be initialized, however for the sake of simplicity we will ignore it in the explanation that follows.

$$y = \begin{cases} 1, & \text{if } o(x) \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Let d represent the *desired* category of the example under consideration (which we know of because as human beings we can identify the true category of the character as either an "A" or a "B"). We can then adjust the weights w_i according to the following learning rule:

$$w_i \leftarrow w_i - \eta(y - d)x_i$$

The motivation behind that rule amounts to:

1. If network output $y = 0$ while target output $d = 1$, then y is too small, so we need to increase the weights associated with positive inputs by a small value η called a *learning rate* while decreasing weights that correspond to negative inputs.

2. Likewise if $y = 0$ and $d = 1$ we need to do the exact opposite so as to lower the value of y.

3. Finally, if the network classified the example correctly (i.e., $y = d$) we do not change anything.

The procedure repeats for several steps, running through the database of images until the average error converges to a minimum.

A similar kind of learning rule was used in the second and fourth interventions of *Absence*. In the *Second Absence* (the light-adjusting system in a bottle), the output itself (i.e., the LED intensity) is directly adjusted using a learning rate of 1 in response to the input (i.e., the photocell measuring the light intensity) (see the algorithm p. 4). In the case of the *Fourth Absence*, an intermediate parameter w is used to control the output (i.e., the rhythm at which a sound is produced); this parameter gets updated in response to the battery voltage using a learning rate η (c.f. page 2.3). In other words, a parameter gets adjusted to generate actions by trying to lower an error rate (measured as the differential between a target and an actual battery charge during nighttime).

Perceptrons mark an important step in the history of Machine Learning for two reasons. These

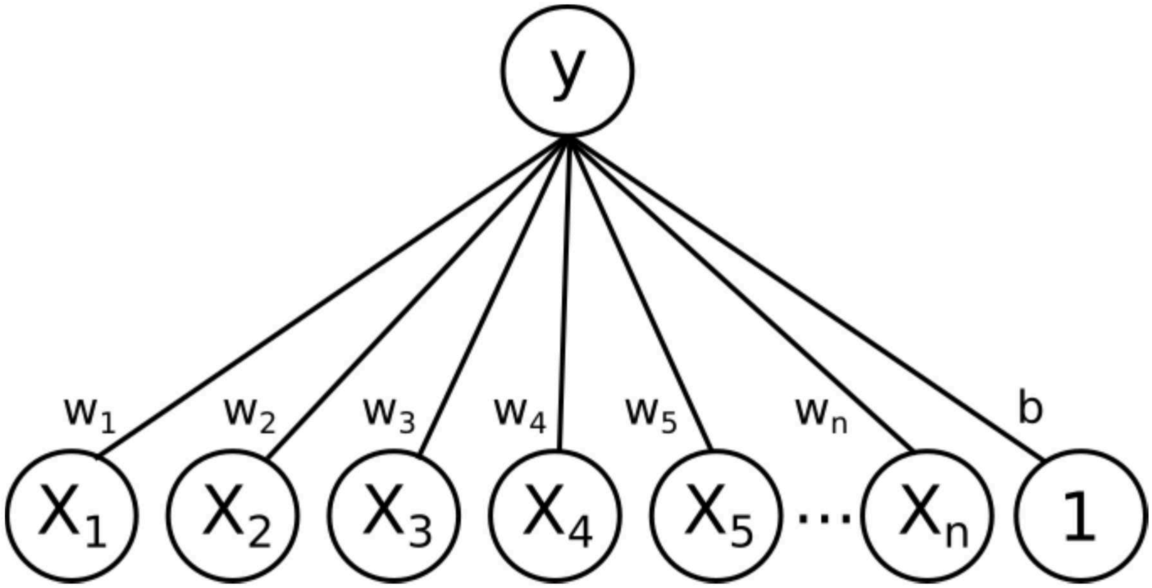


Figure 13: Schematics of the perceptron. Inputs $x_1, \cdots, x_n$ (for example, the value of the pixels of a black-and-white image representing a handwritten character) are multiplied by weights $w_1, \cdots, w_n$ with added bias b to generate output y (which represents a class, for example, whether the image represents letter "A" or "B") .

simple linear systems, which combined ideas about logic, statistics, and self-organization in a computational apparatus that could be used for dedicated purposes through a supervised learning loop, would be used as building blocks for later theories about Machine Learning and Neural Computation. Yet, as we see in the next section, they contain an important flaw that would soon to be revealed, plunging the field into the dark for more than two decades.

3.1.3 Classic AI

Interestingly, Allen Newell, one of the key researchers behind this shift away from connectionist networks, was originally inspired by a self-organizing model similar in spirit to the perceptron. In 1959, Oliver Selfridge proposed the *Pandemonium*, a multi-layered connectionist structure for image pattern recognition (Selfridge 1959). In this structure, simple agents called "demons" are organized in a hierarchy, with the lower-level ones specialized in detecting simple features such as curves and lines, whereas the higher-level ones use the information from the lower layers to detect more complex and abstract features, such as handwritten letters.

The Pandemonium was built on Selfridge's previous research on visual pattern recognition. In 1954, the scientist had given a talk at the RAND Corporation in Santa Monica, describing a system programmed by G. P. Dinneen that was able to learn by experience (Dinneen 1955; Selfridge 1955). Present at the conference was computer scientist Allen Newell who was then conducting research into army-related logistic problems. Newell was deeply inspired by Selfridge's talk. While the learning capabilities of the system were rather poor in practice, it nonetheless revealed a true potential for machines to display intelligent behavior.

In the year that followed, Newell started working on an adaptive system to effectively play chess, which was presented at the Western Joint Computer Conference in 1955. His work grasped the attention of economist Herbert A. Simon at Carnegie Mellon and, together with RAND programmer J. C. Shaw, they started working on the ambitious project of designing a program that would be able to prove mathematical problems.

The program, called the Logic Theorist, was shown the next year during the famous 1956 Dartmouth Conference at Dartmouth College in Hanover, New Hampshire. An initiative of Computer Scientist John McCarthy, the conference brought together a small group of scientists around the study of a new field: artificial intelligence. The study proceeded on the basis of the conjecture that every aspect of learning, or any other feature of intelligence, can in principle be so precisely described that a machine can be made to simulate it.

Whereas several approaches to the problem of computer intelligence were considered, including connectionist methods (McCarthy et al. 2006), the conference was dominated by the work of Newell, Shaw and Simon, who were the only researchers who came with an actual, working artificial intelligence system.

Generally considered to be the first artificial intelligence program, the Logic Theorist was eventually able to prove 38 theorems from Whitehead and Russell's *Principia Mathematica*, even coming up with a more elegant proof for one of them (Newell, Shaw, and Simon 1958; Newell, Shaw, and Simon 1959). It was extremely impressive in its ability to perform in a category of tasks that seemed extremely difficult to humans, requiring a high degree of abstraction and logic.[11] To the contrary of connectionist and Cybernetics approaches, it also did not attempt to model existing biological systems, and instead focused on structured, symbolic manipulations to achieve its remarkable goals.

Newell, Shaw and Simon's work would set the stage for the first phase of the development of the field of AI from the mid–1950s till the mid–1970s. This era was marked, on the one side, by a dubious optimism, as some researchers managed to rapidly achieve satisfying results on high-level problems such as playing checkers or chess (Newell 1955), and responding effectively to simple text-based chat interactions, or solving problems in simulated "micro-worlds" (Winograd 1970); and on the other, by a heavy reliance on symbolic, rule-based systems, with little or no interest in biological systems such as neural networks.

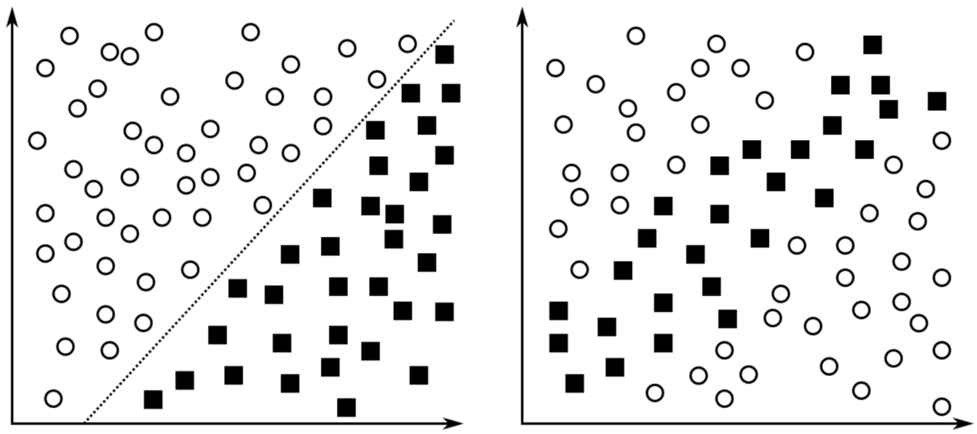
Indeed, the excitement for connectionist structures inspired from human biology that was growing in the 1950s would come to a halt with the publication of Minksy and Papert's forceful critique of perceptrons (Minsky and Papert 1969). By showing that even simple problems are unsolvable by such a linear neural network, the book put a halt to the non-symbolic and distributed approach which had had great attention in the field since the 1940s. Public funding switched sides and for two decades, AI research turned towards the symbolic and heuristic approach pioneered by Minksy, Papert and Simon, which would later be known as "classic AI" or Good Old Fashioned AI (GOFAI).

Classic AI is usually associated with "strong AI" or *computationalism*, a theory of mind based on the premise that cognition is computation (Dietrich 1990).[12] In 1950, Alan Turing proposed a test for machine intelligence using a simple "imitation game". The goal of the machine would be to engage in a continuous chat with a human interrogator and try to "pass" as a human being. If the interrogator could not distinguish the machine from a human interlocutor, then that machine should, according to Turing, be considered as a thinking being (Turing 1950). In other words, what

[11] It would later be found that the most difficult problems for computers to address are not intricate mathematical proofs or efficient strategies for playing board games, but rather the kind of problems that seem so easy to us that we do them almost unconsciously, such as walking, talking, or driving a car.

[12] To understand the importance of this perspective in the history of AI, consider how the preamble of the project proposal for the Darthmouth conference, written in 1955, places it as a foundational component of the field:

> The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. (McCarthy et al. 2006, p. 12)



(a) Linearly separable data. Dotted line represents a possible linear function (such as a perceptron) to split the space in two regions corresponding to each category.

(b) Non-linearly separable data. Linear models such as perceptrons are incapable of correctly separating the space to classify such data points.

Figure 14: Examples of linearly and non-linearly separable datasets in a two dimensional space using two categories of points.

matters when it comes to cognition is not the biological substrate in which it is rooted but the performance of the system on tasks that are deemed to require intelligence.

Hence, computationalists also think that the problem of cognition cannot be solved by mimicking neurological processes using robots or models of the brain, but by using the favored tools of mathematics and computing: logical rules applied to a set of symbols. Its proponents thus believe in the independence of cognition from the "platform" that "runs" it, and also place faith in the Turing test as the ultimate measure of intelligence.

While computationalism is still the dominant viewpoint in the fields of Cognitive Science and AI, it contains many fundamental problems. Philosopher Hubert Dreyfus was one of the first to criticize the computationalist approach of GOFAI. In his 1979 book, *What computers can't do: the limits of artificial intelligence* (Dreyfus 1979), he highlights the false promises of classic AI, that shows a "recurring pattern" of "early, dramatic success followed by sudden unexpected difficulties" (39). Dreyfus attributes this pattern to four inadequate assumptions (biological, psychological, epistemological and ontological) from early AI practitioners. In a phenomenological critique of their approach, he proposes to address these caveats by putting the body back into the equation.[13]

After the impressive results of early AI, research quickly plateaued, plagued by profound theoretical and practical problems (Pfeifer 1996). By the mid–1970s, government support had stopped flowing, leading to a dry period often referred to as the "AI Winter" which ran for about a decade. In the 1980s, new approaches started resurfacing, such as Expert Systems, Artificial Life and Machine Learning. But before we move on, we will have a look at the influence of early conceptions of machine intelligence and adaptation on the artistic culture of the 1960s and 1970s.

Many of my own agent-based installation works before 2008 used some form of rule-based systems that, in the spirit of GOFAI, approached the question of artificial agency through heuristics. For example, my M. A. project *Flag* (2007), an immersive interactive installation, generates sequences of words semantically connected to one another through the construction of a hand-made graph representation of a word cloud. The *First Absence* was also designed using such an approach, using my own *a priori* knowledge of the world to encode a behavioral response (in this case to the sunset). Most agent-based installations that use some form of AI technology actually follow this same approach, often with very good results. Consider for example the excellent work of robotic artists such as Louis-Philippe Demers, Jessica Field, Ken Rinaldo, and Bill Vorn.[14]

3.1.4 Cybernetics and Aesthetics

In the first three decades following the end of the war, Cybernetics, Connectionism and AI offered different perspectives over the functioning of cognition. The apparition of computer-based technologies in society had a tremendous impact in these years. However, how it affected the artistic world is often overlooked. Art historian Edward A. Shanken describes the influence of Cybernetics on art in the 1960s through the work of Roy Ascott (Shanken 2002). Ascott's reading of cyberneticians

[13]There is an extensive body of literature that critiques computationalism. For example, see (Searle 1980), (Winograd and Flores 1987), (Suchman 1987), and (Harnad 1990).

[14]There are many reasons for this, but I will only name a few here. First, the controlled environment of the gallery space offers very few degrees of freedom and it is possible to restrict it even more, thus allowing the effective use of simple computational tricks. In other words: galleries are "small worlds" that can be toyed with in often impressive ways. Second, contrary to AI researchers, artists only need to generate an illusion of agency which does not even need to feel or look smart. Figurative elements are really important in the production of a sense of agency. Picture how drawing two eyes and a mouth on an inanimate object such as a potato can suddenly transform it into an agent, at least perceptually. Finally, the public itself is adaptive and, in an artistic setting, often ready to "suspend their disbelief" as long as they are going to have a good show.

Norbert Wiener, Ross Ashby and Frank H. George in 1961 made him envision a new conception of art as something embodied in interactive systems rather than in physical objects. The scope of Cybernetics as an encompassing theory of systems' behavior and communication, would allow Ascott to merge Cybernetics and art, in an effort to "theorize the relationship between art and society in terms of the interactive flow of information and behavior through a network of interconnected processes and systems" (4).

Cybernetics' conceptions of adaptivity, homeostasis and feedback loops are thus an integral component of Ascott's perspective, which he explains in his 1966 paper *Behaviourist Art and Cybernetic Vision* (Ascott 2003a). In it, he claims that visual arts have entered a new era where other modalities (such as sound and touch) are explored by artists and where the interactive and participative experience of the spectator in relationship with the artwork becomes central. Ascott thus suggests the name "behavioural art"[15] as a replacement for "visual art", which has become too narrow to describe the new paradigm (110). Ascott later argues more specifically what he means by it and its relationship with adaptivity and feedback:

> Behaviourist art constitutes [. . .] a retroactive process of human involvement, in which the artifact functions as both matrix and catalyst. As matrix, it is the substance between two sets of behaviours; it exists neither for itself nor by itself. As a catalyst, it triggers changes in the spectator's total behaviour. Its structure must be adaptive, implicitly or physically, to accommodate the spectator's responses, in order that the creative evolution of form and idea may take place. The basic principle is feedback. The artifact/observer system furnishes its own controlling energy: a function of an output variable (observer's response) is to act as an input variable, which introduces more variety into the system and leads to more variety in the output (observer's experience). (128)[16]

Hungarian artist Nicolas Schöffer's piece *CYSP I*, created in 1956, is considered to be the first autonomous cybernetic sculpture to follow Ascot's definition.[17] The work consists of an eight-foot

[15] Ascott seems to use the terms "behavioral" and "behaviourist" interchangeably in his writings.

[16] Roy Ascott would enact this vision not only through his art practice but also through his views on the pedagogy of art. In his 1967 manifesto Behaviourables and Futuribles, Ascott frantically argues for restructuring art schools "as homeostatic organisms, living, adaptive instruments for generating creative thought and action." (Ascott 2003b, p. 159)

[17] In fact, the name of the work comes from a combination of the two first letters of cybernetics and of Schöffer's theory of spatiodynamic sculpture.

tall metallic structure that can move freely across space thanks to four rollers set at its base. At the top of the sculpture are mounted sixteen (16) motorized plates of colored acrylic glass. A central processing unit designed by the Philips corporation uses the signals coming from a set of photo-electric cells and a microphone to control the motors, allowing the device to respond to its environment and to itself, thus engaging in a self-organized behavior. This design allows it to adapt to different contexts of presentation: shown for the first time in Paris as part of “Nuit de la poésie”, it would later join human dancers in a *pas de deux* by choreographer Maurice Béjart and, in 1957, would participate in a “spectacle cybernétique” in Évreux under the musical direction of Pierre Henry (Schöffer 2004). Directly inspired by Norbert Wiener’s theory of control and communication, Schöffer’s work is a pioneering example of the kind of feedback systems early cyberneticians had in mind (Fernández 2006, 472).

British polymath Andrew Speedie Gordon Pask is another key figure of the cybernetic art movement which was lead by Roy Ascott in the UK.[18] Pask had allegedly discovered Cybernetics as an undergraduate at Cambridge in the early 1950s, through an impromptu meeting with Norbert Wiener himself (Pickering 2010, 313). While he his mostly known for his scientific work, Pask’s involvement with Cybernetics first started in the art world. During his years at Cambridge, Pask had participated in the lighting design of theatrical shows in Cambridge and London and created, together with fellow student Robin McKinnon Wood, a business specialized in the orchestration of musical comedies. In 1953, they invented a theatrical lighting system called the *Musicolour*, an apparatus that “used the sound of a musical performance to control a light show, with the aim of achieving a synesthetic combination of sounds and light”(316). Reacting adaptively to a sound signal, it generated patterns of light, interacting with human performers in real-time. The device contained a “rudimentary learning facility” that was able to change the relationship between sound and light during the course of a performance.

Notions of adaptation and learning are what fascinated Pask the most in cybernetics systems. Discussing the evolution of the work as it toured across the country, he notes:

[18]Evidently, it is Pask who originally explained cybernetics to Ascott at his request (Miller 2014). Pask and Ascott would actually get to work together in the early 1960s as consultants on Price and Littlewood’s Fun Palace, an ambitious Cybernetics architectural project that would never be built (Mathews 2005).

> By that time it was clear that the interesting thing about Musicolour was not synaesthesia but the learning capability of the machine. Given a suitable design and a happy choice of visual vocabulary, the performer (being influenced by the visual display) could become involved in a close participant interaction with the system. He trained the machine and it played a game with him. In this sense, the system acted as an extension of the performer with which he could co-operate to achieve effects that he could not achieve on his own. Consequently, the learning mechanism was extended and the machine itself became reformulated as a game player capable of habituating at several levels, to the performer's gambits. (Pask 1971, 78)

Along with artists Nam June Paik, Jean Tinguely, John Cage and Edward Ihnatowicz, Pask would participate in 1968 in the exhibition "Cybernetic Serendipity" at the Institute of Contemporary Arts in London, with his work *Colloquy of Mobiles*, which involved a social and sexual metaphor of agents trying to collectively achieve a goal by adapting to one another (Reichardt 1968). The exhibition, curated by Jasia Reichardt, set a landmark in the historical upbringing of media art. It brought together the work of more than a hundred contributors, the majority of which were not artists, such as computer scientists, engineers, and philosophers, as well as a chaotic mix of apparatuses and installations that were purposely intended to confuse the visitor as to whether they were created by an artist or a scientist. Many of the works displayed were, like Pask's installation, directly inspired by Cybernetics, using principles of feedback as their core mechanism (*Cybernetics Serendipity - Late Night Lineup* 1968).

The same year, artist and critic Jack Burnham published "Systems Aesthetics" in *Artforum*, where he explained how the society of the time, shaken by the rapid progress of science and technology, was transiting from an "object-oriented to a systems-oriented culture" where "change emanates, not from things, but from the way things are done." (Burnham 1968, 31). This transition, he claims, is reflected in contemporary practices emerging in the 1960s such as Robert Smithson's "earthworks", the light "sculptures" of Dan Flavin, the "kinetic art" of Jean Tinguely and Alexander Calder as well as Allan Kaprow's Happenings.

Burnham's perspective over art and science is intimately linked with conceptual art, which in the 1960s promulgated the supremacy of ideas over forms. Indeed, Burnham's epitomic 1970 exhibition *Software* brought together conceptual artists such as Vito Acconci, John Baldessari, Robert

Barry, Donald Burgy, Hans Haacke, Douglas Huebler and Joseph Kosuth, to explore computer technologies as ways to generate interactions between the audience and machines. "Systems aesthetics" established a link between Grey Walter's creative experiments with mobile robots and the work of pioneering cybernetic artists such as Nam June Paik and Nicholas Schöffer through their interest in imitating life.

Whereas Burnham's visionary perspective was directly influenced by Cybernetics, the concept of systems on which it was built originates from General Systems Theory, an interdisciplinary approach originally formulated by biologist Ludwig von Bertalanffy in the 1930s as well as in different publications after the war. Observing that different disciplines were in fact dealing with similar problems, he argued for an integrated approach that could be applied across them:

> Thus, there exist models, principles, and laws that apply to generalized systems or their subclasses, irrespective of their particular kind, the nature of their component elements, and the relations or "forces" between them. It seems legitimate to ask for a theory, not of systems of a more or less special kind, but of universal principles applying to systems in general. (Bertalanffy 1969, 32)

General Systems Theory and Cybernetics are very close in spirit and in practice. In fact, as pointed out by interdisciplinary researcher Francis Heylighen at the Free University of Brussels, both approaches focus on the same problem of "organization independent of the substrate in which it is embodied" using only a slightly different approach. "[S]ystems theory has focused more on the structure of systems and their models, whereas cybernetics has focused more on how systems function". They should be considered as two faces of the same coin (Heylighen 2000, 460—461).

Burnham argued how art as an institution could be understood as a hierarchical system, with artists as its basis being "similar to programs and subroutines", with, at the very top, a "metaprogram" that constantly rearranges the long-term objectives of art. Key to Burnham's vision is the conclusion that this self-organizing, adaptive system does not produce new objects, but rather new information embodied in works of art (Burnham 1969). In 1970, he curated the show *Software* at the Jewish Museum in New York, where he articulated this vision by bringing together works created by artists and scientists alike that made extensive use of computer technology, with the objective of generating aesthetic effects without the intervention of objects. "The machines in Software", he claimed, “should not be regarded as art objects; instead they are merely transducers, that is, means of relaying information which may or may not have relevance to art.” (*Software* 1970, 12)

Burnham’s systems art and Ascott’s behaviourist art both translated ideas about the adaptive and emergent nature of human and nonhuman systems inherited from scientific research in Cybernetics and General Systems Theory. While both Burnham and Ascott have been relatively overlooked by art historians and theorists, their aesthetics have ramifications in many art forms from the 1960s onward, such as conceptual art, information art, algorithmic art, generative art and robotic art.

Yet, when applied to works such as *Absences*, both Burnham and Ascott’s theories have shortcomings. One of the strengths of Burnham’s framework is to relocate the locus of aesthetics from the physical properties of an art object into the artwork’s mode of operation as a system embedded in a network of relationships. Burnham’s emphasis on a disembodied flux of information that channels through the art object — which only act as an empty shell independent from the process that it allows to run — is, however, reminiscent of computationalism as a model for the workings of the brain. Burnham’s vision is thus somehow tainted by the Shannionian myth of aerial, disembodied processes that run independently from their corporeal substrate. This argument is problematic when applied to an aesthetics of agent-based systems as it fails to take into account the question of the embodiment of these agents.

Ascott, on the other hand, seems to be interested in taking into account not only the production of novelty as pure information but also in its morphological evolution. His perspective, however, presupposes an interaction between a work of art and a human, which seems less appropriate in the case of the nonhuman-to-nonhuman dynamics that happen in *Absences*.

Still, both Ascott and Burnham highlight an important point in their focus on *behaviors* — and the experience of such behaviors — in machine-human configurations in the artistic domain. Furthermore, their interest in Cybernetics aligns with a vision of society, culture, and art, as profoundly adaptive systems, evolving through a network of self-organizing agents which adjust to one another through a myriad of feedback loops.

3.1.5 Machine Learning

At the beginning of the 1980s, classic approaches in AI were still dominant, showing no interest in any form of biologically-based computation such as genetic algorithms and neural computation. Nevertheless, a small portion of AI researchers had become interested in questions of learning systems. Pat Langley describes the creation of the new discipline of Machine Learning in the early 1980s:

> The first workshop in 1980 at Carnegie-Mellon University had identified a community of researchers with common interests in computational approaches to learning and arrived at a name for its activities. Moreover, the parent fields of artificial intelligence and cognitive science were showing little interest at the time in learning-related issues, preferring to focus on the role of knowledge in intelligence, regardless of its origin. As a result, we encountered some difficulty publishing and, more generally, felt we were not getting the attention we deserved. Finally, there was the common urge of young, energetic researchers to create something of their own to which they could attach their names. (Langley 2011, 275)

Machine Learning is a sub-field of Artificial Intelligence that employs mathematical models that can classify and make predictions based on statistical inference over observed data rather than on logical rules. It can be split in three main areas of inquiry: supervised learning, unsupervised learning and reinforcement learning. Supervised and unsupervised learning methods are used for statistical classification or regression of data points.[19] Supervised learning is used when we know in advance the target category of the data points we want to classify, such as when trying to recognize hand-written digits, whereas unsupervised learning is when we do not have tagged data points but rather want to learn some inherent properties of the data distribution under consideration. Finally, reinforcement learning (RL) rather tries to address the problem of an agent adapting to its environment by trying to optimize a criterion called a reward function, which basically rewards or punishes the agent depending on its current state and actions (Sutton and Barto 1998).

[19] Classification consists in assigning a label or class to a data point, for example the symbol (A, B, C, etc.) represented by a handwritten letter, or the name of an individual recognized in a photograph. Regression, on the other hand, concerns the estimation of a numerical value, for example, the expected amount of claimed dollars that will be made by an insurance customer, or the expected temperature for tomorrow.

3.1.6 Connectionism in the 1980s

At its beginning, the new field of Machine Learning was still mostly based on symbolic methods such as decision trees and logic. But in the middle of the decade, the discovery of an efficient way to train Multi-Layer Perceptrons (MLP) would suddenly bring connectionism back on the scene (Rumelhart, Hinton, and Williams 1986). MLPs — which are also often called Feedforward Artificial Neural Networks (FANN) — can be used for classification as well as for regression (i.e., function approximation). As their name indicates, they consists in stacking many perceptrons on top of each other in interconnected layers of neurons. Hence, they differ from perceptrons in that they not only have an input and an output layer of neurons but also one or more hidden layers between these inputs and outputs. Like in the perceptron, a first set of weights maps the input neurons to an intermediate "hidden" layer, where abstract, higher-level representations of the inputs are automatically generated through the cooperation of neurons.

One way to understand these structures is to consider each hidden neuron as the output neuron of a perceptron. The difference is that in the case of the perceptron, the output is transformed into a binary value using a hard threshold. In a MLP, the hidden neurons are transformed using a smooth, non-linear thresholding function that pushes them towards a binary value. Finally, the hidden neurons are linearly combined using a second set of weights to produce the next layer of neurons.[20] This process moves forward, from layer to layer, until the final output layer is reached, yielding the result.

Because each layer projects the previous layer's outputs using a non-linear thresholding function,[21] MLPs model smooth classification functions that can grasp intricate, high-order variations

[20] One can look at an ANN as a network of agents, where each hidden neuron is seen as a minimal agent that becomes an expert classifier over a specific domain. These agents are encouraged to divide the input space between them. They are then combined to produce the final output, as if they were "voting".

[21] As can be recalled from section 3.1.2, perceptrons are simple linear models that separate space using a hyperplane. They are thus incapable of dealing with problems that are not linearly separable, as was rightfully pointed out by Minsky and Papert (Minsky and Papert 1969). Indeed, most interesting problems turn out to be non-linearly separable.

MLPs' "tour de force" consists in preserving the self-organizing, distributed representation properties of perceptrons, while alleviating their flaws by applying a non-linear filter to their outputs. A very commonly used filter is the sigmoid function, which was described in section 2.3 .

in the data, thus circumventing the main caveat of perceptrons as pointed to by Minsky and Papert (Minsky and Papert 1969). But while MLPs were actually known way before the mid–1980s[22], there existed no tractable way to train them. The 1986 breakthrough, introduced by David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, consisted in a method known as *backpropagation* that allowed the efficient computation of the neural network's weights' gradient.[23] The gradient essentially represents the partial derivatives (in other words, the degree of change, or the slope) of the error function with respect to the weights. Knowing it thus gives a sense of how the error changes in function of the weights, allowing the adjustment of weights in a stepwise manner in the direction that is most likely to lower the error, a procedure known as Stochastic Gradient Descent (SGD).[24]

Without entering into details, consider the similarity of the update rule in SGD to that of the perceptron. Let E be the error function, w be a weight in the MLP, and $\frac{\partial E}{\partial w}$ be the partial derivative of the error with respect to that weight. After each step of SGD, the weight will be changed using the following formula:

$w \leftarrow w - \eta \frac{\partial E}{\partial w}$

Once again, we retrieve a negative feedback procedure that fine-tunes a weight by pushing it in the direction that will enable it to most likely contribute to better classifications (i.e., lower errors) in the future, tempered with an adjustable learning rate η that controls the speed of learning.

Like real neural networks in the brain, ANNs represent information in a distributed way, as opposed to a symbolic, local representation. At the beginning of the procedure, the weights are initialized randomly, such that the network decisions are completely chaotic (i.e., the entropy is maximal). By getting exposed to the environment (in other words, by being subjected to examples sampled from the real world distribution) and taking actions in a range of different contexts, the network is slowly adjusted to make better predictions. Thus, the network itself becomes increasingly ordered as its parameters (weights) are shaped to decrease the global entropy of the model.

[22]Selfridge's Pandemonium, introduced earlier, grasped the concept of interconnected layers of abstractions already in the 1950s (Selfridge 1959).

[23]In fact, backpropagation had been discovered years before but had not been applied specifically to neural networks. For a detailed historical account of backpropagation, read (Schmidhuber 2015).

[24]As a metaphor, imagine a ball rolling down a hill, always going in the steepest direction until it reaches a minimum.

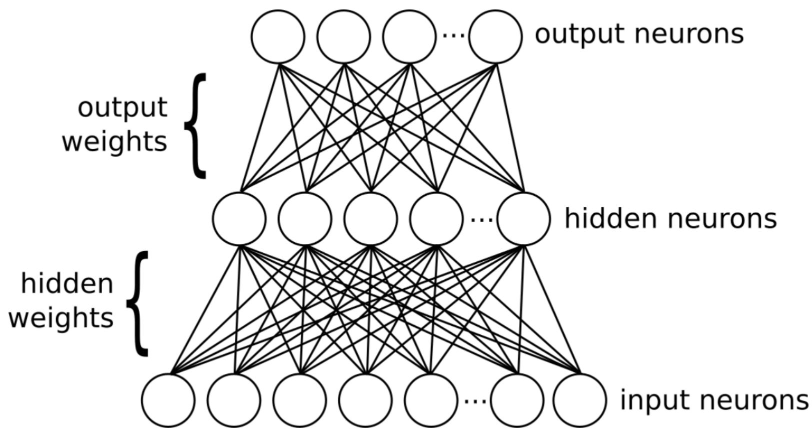


Figure 15: Schematics of a Multi-Layer Perceptron with a single hidden layer. The first layer is a perceptron that maps input neurons to hidden neurons that act as intermediate outputs. The second layer is another perceptron that uses these hidden neurons as inputs and maps them to the final output neuron of the model.

The revival of connectionist adaptive systems in the 1980s had a tremendous impact on the emergence and development of Machine Learning and neural computation as fields of research. In 1987, the first Conference on Neural Information Processing Systems (NIPS) took place, bringing together researchers interested in connectionist approaches from both neurosciences and computer science and would become, over the years, the most important conference in the field of Machine Learning. The same year, Stephen Grossberg launched the International Neural Network Society (INNS) as well as its associated publication, *Neural Networks.*

Recalling a conversation with Grossberg in 1987 where he was asked to join the INNS and become a co-editor of the journal, Finnish scientist Teuvo Kohonen gives an insight into the important connections between neural systems and AI at the origins of Machine Learning in the mid–1980s:

> Then he started to talk about the term "neural". I said, "No, no, no, no." I said, "Why not learning machines or adaptive systems or whatsoever?" So he said, "Yes, but we have so many opinions, and this seems to satisfy everybody." (Anderson and Rosenfeld 1998, 153)

3.1.7 Evolutionary Computation

Genetic Algorithms (GA) is another approach to Machine Learning that was largely popularized in the 1980s. GAs stemmed from a completely different branch than neurology: that of genetics and evolution. Although developing through its very own path, it is important to describe it here, mainly because it is one of the learning approaches that has been the most largely adopted by artists.[25]

While the were many research groups in the 1960s working on computational models of evolution applied to AI,[26] the invention of Genetic Algorithms (which is one of many approaches to evolutionary computation) is usually attributed to scientist John Holland (Mitchell 1998), who developed them in an effort to build a formal mathematical representation of genetic adaptation that could be run on computerized systems. Holland's framework understands natural evolution as an iterative optimization process that functions by evolving populations of individuals using basic genetic operators such as cross-overs and mutations, testing them against a fitness function[27] and selecting only the best individuals to generate the next population (Holland 1992). The basic form of GAs as proposed by Holland employs artificial chromosomes that are essentially sequences of bits (i.e., the genotype). Segments of the string correspond to genes that determine actual characteristics of the individual (i.e., its phenotype). The performance of the individual can than be assessed using a fitness function which determines what needs to be optimized.

A genetic algorithm in its simplest form goes more or less as follow (Mitchell 1995, 5):

1. Begin with an initial population of N individuals (i.e., chromosomes).

2. Select the M fittest individuals according to fitness function $F(x)$.

3. Perform crossovers and mutations over pairs of selected individuals, thus generating a new

[25]The reasons for this are unclear. However, one can point to the fact that GAs are more easy to understand, implement and apply in works of art, as one possible explanation.

[26]For example, see Rechenberg's *evolutionsstrategie* (Rechenberg, 1965/August//; Rechenberg 1973) as well the "evolutionary programming" by described in Fogel, Owens, and Walsh (1967).

[27]A fitness function is an evaluation function that gives a value (typically as a real number) to an individual in a population, usually representing its performance over a problem that the GA tries to solve.

For example, a GA used to learn how to play chess might generate populations of agents and have them play games against one another. The fitness function could then be the percentage of games won (with ties counting as a half-win).

generation of N offsprings.

4. Replace the population with the newly generated one.

5. Go to step 2.

This is an optimization algorithm: it performs a search through the space of possible solutions to a problem, represented as computational chromosomes, using an evolutionary heuristic. The process works by making local changes in potential solutions to a problem, moving closer to the end goal in a stepwise manner. It is thus very close, in essence, to other adaptive algorithms such as Stochastic Gradient Descent.

3.1.8 Adaptation and Learning

This historical overview exposes the important role played by adaptation and learning in the scientific landscape in regards to computational cognition and machinic life, from the post-war era onward. Yet, these ideas and associated techniques seem to have only been rarely exploited directly by artists and even less examined by humanities scholars. Before we move into the second part of this chapter, I would like to discuss further these notions by considering different scientific definitions.

As suggested earlier, we can trace the origins of modern understandings of adaptivity down to first-order Cybernetics in the work of Arturo Rosenblueth, Norbert Wiener and Julian Bigelow (Rosenblueth, Wiener, and Bigelow 1943). Core to their conception of system processes is the *teleological* nature of animal behavior, that is, their ability to adjust to their environment to reach their goals using negative feedback. This fundamental idea constitutes the cornerstone of Machine Learning approaches, which requires the interaction between a structural component, an error function and an optimization procedure, as I will further explain in section 3.2.

The authors suggest a hierarchical taxonomy of behavior. In that ontology, living systems are first said to be *active*, in that they are the source of energy that allows their actions, as opposed to passive objects like a stone being thrown by another agent. They are also considered *purposeful*, which relates to their behavior appearing as being directed voluntarily towards a goal. Some

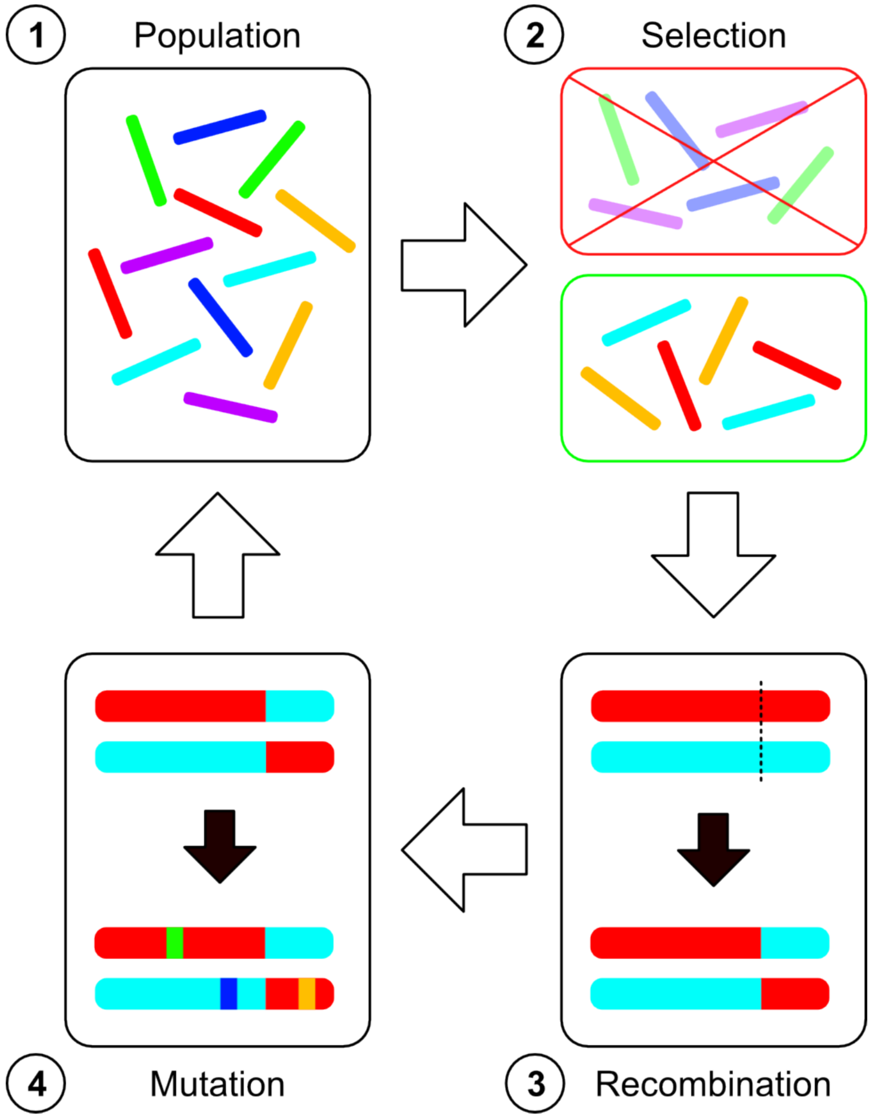


Figure 16: Schematics of a genetic algorithm. (1) A population of genes is translated into a population of phenotypes. (2) These phenotypes are evaluated using a fitness function: the fittest individuals are selected. (3) Crossovers and (4) mutations are performed over the genes of the selected individuals, thus generating a new population of genes.

purposeful active behaviors are also *teleological*, meaning that they use *negative feedback* to adjust their aim while trying to reach their goal.[28]

British cyberneticians like Ross Ashby, Grey Walter, Gordon Pask and Stafford Beer proposed visions of Cybernetics systems that can be seen as alternatives to these behavioral categories. These researchers seemed less interested in a designing a theory of final causes than in imagining and articulating ways in which living systems perform, in particular through the creation of devices endowed with lifelike qualities. Discussing their work, Andrew Pickering writes:

> There is something strange and striking about adaptive mechanisms. Most of the examples of engineering that come to mind are not adaptive. Bridges and buildings, lathes and power presses, cars, televisions, computers, are all designed to be indifferent to their environment, to withstand fluctuations, not to adapt to them. The best bridge is one that just stands there, whatever the weather. Cybernetic devices, in contrast, explicitly aimed to be sensitive and responsive to changes in the world around them, and this endowed them with a disconcerting, quasi-magical, disturbingly lifelike quality. (7)

Grey Walter created his electro-mechanical "tortoises" during his spare time. His later versions testify to an interest in a machinic form of learning inspired from behaviorism, even though he does not explicitly use the expression "Machine Learning" to describe their behavior (Walter 1951). The first alleged use of the term comes from engineer Arthur Lee Samuel who worked — also in his time off — on the game of checkers. Samuel was interested in achieving better results at playing the game not through logical rules or brute-force search, but rather by providing an algorithm with several instances of played games, allowing the system to learn by itself what are the best moves. His technique, which he describes in his 1959 foundational paper (Samuel 1959), can be considered as an embryonic instance of reinforcement learning (Sutton and Barto 1998, 267).

In the preface to his foundational book *Adaptation in Natural and Artificial Systems*, John H. Holland proposes a formal definition of adaptation as a "process whereby a structure is progressively modified to give better performance in its environment" (Holland 1992, xiii). In this perspective,

[28] To understand the difference between teleological and non-teleological systems, consider the example of a snake striking at a frog with "no visual or other report from the prey after the movement has started". As the authors express, in that case the movement is "so fast that it is not likely that nerve impulses would have time to arise at the retina, travel to the central nervous system and set up further impulses which would reach the muscles in time to modify the movement effectively.

the story of evolution can be seen as an optimization process that performs a heuristic search in the realm of possibilities by selecting the best individuals at each generation, preserving part of their genetic structure while combining and mutating them. Significantly, the same kind of genetic procedure which nature applies to evolve fittest forms or organs in living systems can be digitally simulated to develop better strategies of action in computational agents:

> Roughly, experience guides changes in the organism's structure so that as time passes the organism makes better use of its environment for its own end. (Holland 1996, 9)

As a final remark, to better contextualize the place of adaptation and learning in post-WWII discourses on artificial intelligence, it is worth stating that agent-based approaches to learning and Genetic Algorithms were substantially sidestepped during the connectionist renaissance of the mid–1980s — which established the foundations of Machine Learning as a field of research in its own rights — in favor of pattern recognition applications.[29] The neural approach to learning decreased in popularity throughout the 1990s in favor of a more general conception of computational adaptation known as probabilistic or statistical learning. Gaussian Mixture Models (GMMs) and Support Vector Machines (SVM) are examples of such approaches that rely primarily on statistics rather than on some model of biological processes (Vapnik 2000).

3.1.9 Deep Learning

What mainly explains the decrease in popularity of connectionism as an approach to AI in the 1990s was the problem of training artificial neural networks with many layers of neurons. This prevented such systems to grasp highly varying functions, which are needed to express complex behaviors such as the ones we find in "intelligent" agents such as humans.

Until the mid–2000s, it was only possible to train *shallow* neural architectures efficiently — that is, connectionist networks with only 1, 2, or 3 layers. But many neuroscientists seem to believe that brains are organized in *deep* architectures, processing sensory information through many different levels of abstraction (Serre et al. 2007; Bengio 2009). For example, the visual cortex contains

[29] However, GAs were very important in artificial life and evolutionary art communities.

multiple layers of neurons that correspond to different degrees of representation, from detecting edges and orientations to more complex shape recognition (Kruger et al. 2013).

In the early 2000s, raw computational power became more readily available to scientists, fostering what Machine Learning expert Jürgen Schmidhuber called a "second Neural Network ReNNaissance" (Schmidhuber et al. 2011) — a reference to their "First Renaissance" in the 1980s triggered by the publication of the backpropagation algorithm by David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams. This allowed researchers to run experiments over much larger models, encouraging the development of algorithmic techniques to address the shortcomings of shallow architectures.

In 2006, Geoffrey E. Hinton (who was part responsible for the aforementioned first "Renaissance" of neural nets), Simon Osindero, and Yee Whye Teh came up with a solution for training Deep Belief Networks, which are a special kind of multi-layered neural network. This development plugged the breach left open twenty years ago after Hinton last published his work on backpropagation (Hinton, Osindero, and Teh 2006).[30] The method they proposed used unsupervised learning to pre-train the lower layers of the model before subjecting the whole system to a traditional supervised learning procedure. Their approach created a significant improvement in the error rate of MNIST (a benchmark database well-known in the field of Machine Learning) over other approaches using shallow architectures or Support Vector Machines.

This breakthrough, along with much other research, allowed for the emergence of a whole new field within Machine Learning called Deep Learning, whose main interest lies in finding solutions to difficult problems (such as driving a car) by allowing computers to "learn from experience and understand the world in terms of a hierarchy of concepts, with each concept defined in terms of its relation to simpler concepts". This system therefore avoids the need for humans to "formally specify all of the knowledge that the computer needs". In particular, this "hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones" (Goodfellow, Bengio, and Courville 2016).[31]

[30] This is an oversimplification of the history, of course, as scientific discoveries do not happen in a vaccuum. For a more thorough analysis of the history of deep learning, please consult (Schmidhuber 2015).

[31] Deep Learning models have become extremely complex and powerful, with some neural nets being more than

While none of the artworks described in this dissertation make use of deep architectures, I wish to discuss Deep Learning here because of its important socio-technical repercussions. As mentioned in the introductory chapter, its successes have now made Deep Learning the spearhead of big IT companies such as Google and Facebook, who have become the main hiring bodies for Deep Learning engineers. After the media success of its Deep Dream project, Google has recently launched Magenta, a research project that seeks to advance and explore the applications of AI relative to "music and art generation" (Eck 2016).[32]

While Google's interest in art and creativity sounds like good news to media art professionals, a word of caution is necessary. One should not forget that the multi-billion dollars companies' main source of profit is advertising: in other words, their main use of Machine Learning and Data Mining algorithms is to better target consumers who freely give their data through the use of their platforms. I do not place faith in techniques developed by such companies, as I doubt their capability to be particularly revolutionary or challenging aesthetically if they are intended to appeal to a mass consumer market. I also notice that there are severe power unbalances between artists and these companies, often resulting in a situation where artists are effectively functioning as an underpaid cultural currency at best, and underpaid technological labor at worst. As such, I and other cultural critics see art and science initiatives such as the Google Cultural Institute and the Facebook Artist in Residence program as branding strategies that also offer convenient ways to capitalize on artists, by getting access to their ideas and expertise at a low cost.[33]

3.2 Machine Learning in Art Practice

Cybernetics-style adaptive systems have evolved from the 1980s onward into the science of Machine Learning, bringing together a vast multitude of approaches ranging from statistics, stochastics and Bayesian logic to neural and genetic computing under a common research program within AI. Machine Learning explores algorithms that are able to make inferences and predictions about the

1000-layers deep. . .

[32]It is unclear what they mean by "art" (as opposed to "music") in this description.

[33]Read (Wilk 2016) for an in-depth analysis of the problematics raised by such initiatives.

world by looking at large quantities of data.

Clearly, these techniques were never intended to be used for artmaking. Artificial Intelligence in general, and Machine Learning in particular, have only recently been applied to artistic creation, being traditionally focused on rational problem solving and optimization (Eigenfeldt, Burnett, and Pasquier 2012; Mateas 2001). Art usually attempts to ask more questions than it tries to solve, and does not provide the kind of objective criteria one needs to perform optimization.[34] Still, the excitement and fascination one feels while observing an agent tentatively trying to achieve the arduous task of balancing a pole (Sutton and Barto 1998), performing acrobatic stunts with a toy helicopter (Ng et al. 2004; Abbeel et al. 2007) or finding new ways to play Pong (Mnih et al. 2013), show these experiences possess expressive and aesthetic potentials. But what are the dimensions of Machine Learning algorithms that can be exploited for artistic expression, and how? As a way to approach this question, let us examine the fundamental characteristics of learning methods and explore ways they can be harnessed for art creation.

A Machine Learning algorithm comprises four components: (1) the *category of task* one is trying to solve; (2) the *model* used to address it; (3) the *loss function* against which the model is trained; and (4) the *search or optimization* procedure. These items represent interdependent dimensions of a learning system which come to influence its outcomes — in particular, its aesthetic potentialities.

3.2.1 Tasks

As explained earlier, the field of Machine Learning is divided in three sub-fields, corresponding to three different classes of problem: (1) supervised learning; (2) unsupervised learning; and (3) reinforcement learning. These categories do not exist in utter isolation. Quite permeable, they often share models and algorithms, as the research carried out in one domain can often be applied to another. One famous example of this is the so-called "deep learning" breakthrough which involved unsupervised learning as a key component in training neural network architectures with several layers of neurons on both supervised and reinforcement learning tasks (Hinton, Osindero, and Teh

[34] In other words, it is not clear at all what one would try to optimize. There is no such thing as an "optimal" work of art.

2006).

Supervised Learning

Supervised learning — the most common category — concerns the problem of predicting an output associated with a certain input data, based on a dataset containing examples of data points with expected target response (typically hand-labelled by humans). Two sub-cases exist: (1) *classification*, which consists in determining the correct category of a data point; and (2) *regression*, where a continuous value needs to be predicted. A typical classification task is pattern recognition, for example recognizing hand-written digits. In this case, input data (pixel values) is labeled by humans with a corresponding category (the digit that was written). Using only these examples, the algorithm needs to learn how to recognize unseen examples correctly.

Most of contemporary applications of Machine Learning are supervised learning tasks, such as speech recognition, spam detection, face recognition, and medical diagnosis. There has also been much research carried on their artistic use in the past two decades, with often impressive successes, such as in the field of music generation (Eck and Schmidhuber 2002; Boulanger-lewandowski, Bengio, and Vincent 2012). Since supervised learning can be used to estimate probability distributions, it is possible to train models such as ANNs on a database made up of all of Chopin's work: the resulting network could then be used to randomly generate a score that would sound like a Chopin piece.

Similar experiments have been done in the visual arts for generating images. Australian engineer and digital artist Jonathan McCabe has created a piece called *Nervous States* (2006) consisting in a series of prints generated by neural nets. The images come to reveal the underlying organization of the system:

> The X and Y coordinates correspond to two variables in the connections of the network; the colour of the pixel at that point is a representation of the network's behaviour for those parameters. So the image is a map of system states; coherent colours show areas of relative stability or gradual change; edges show sharp jumps in the output; marbled swirls show complex oscillations. (Whitelaw, Wednesday, August 16, 2006)[35]

[35] Another research project analogous to this one that has gained a much wider audience is Google'd DeepDream program, which produces psychedelic images based on the response of one of their deep neural network architectures which is fed back into itself (Mordvintsev, Olah, and Tyka 2015).

Rebecca Fiebrink, a computer scientist from Goldsmiths University, has created the *Wekinator*, an open-source software that employs supervised learning as a tool for music performers and artists. The software allows practitioners to easily train ML models to recognize gestures and map them into any kind of contents. One advantage of the approach is that it opens up the possibility for non-coders to create intricate relationships between movements and content in real-time.

An important work from the mid–1990s, Karl Sims' *Galápagos* (1997), uses a GA trained in real-time using supervised learning. The piece is representative of his interest in the evolution of morphologies and behaviors. A series of twelve computers each show a single virtual 3D organism whose shape and movements are the phenotypic outcome of a digital genotype. Visitors interact with the piece by choosing which organisms they prefer. The selected individuals are then used to generate a new population of organisms using a Genetic Algorithm that mates, recombines and mutates their digital DNA, producing offsprings that resemble their parents (Sims 1991, 1994).

Galápagos rests on Sims' astute application of a GA process where visitors take part in a retro-action loop by evaluating the fitness of virtual creatures according to their (subjective) aesthetic qualities. The artificial life forms are thus evolved so as to adapt to the audience's tastes over the exhibition period.

Supervised learning has thus been used for creative expression in generative art and performance-driven works. In agent-based art that relies on the design of real-time behaviors, such as robotic installations, it has mostly been utilized in less experimental manners, such as for computer vision, often applying off-the-shelf solutions. These works are less interesting for the current study, as the fact that they are based on learning does not have a strong impact on the resulting behaviors produced. In other words, these works might simply perform better than previous non-learning systems, but it does not change their final aesthetics in a significant manner.

Unsupervised Learning

Unsupervised learning refers to classes of problems wherein there are no precise outputs that need to be predicted, typically referred to as "unlabeled data". Rather, the algorithm needs to learn "something about the data distribution". Tasks include (1) *clustering*, where the system is asked to split the data space in different regions, or classes; (2) *dimensionality reduction*, where it tries to extract the most important regularities in a distribution to represent it using less dimensions; and (3) *representation learning*, where the model attempts to learn “good” representations of the data, usually to be fed as inputs to other Machine Learning systems.

Unsupervised learning techniques are particularly interesting for artistic research and production, as they give more space for the learning models to come up with potentially surprising solutions, whereas supervised learning methods aim to achieve clear, definite goals in the most accurate way possible; a property that is certainly useful for engineering applications, but quite restrictive in its ability to generate novel content.

Indeed, one of the learning methods that has been the most widely used by new media artists is an unsupervised neural network called a Self-Organizing Map (SOM). SOMs were invented in 1981 by Teuvo Kohonen — they are often called Kohonen Maps — and would become one the most famous unsupervised learning techniques (Kohonen 1981, 2001).[36] They can be thought of as a kind of Perceptron that can be trained to learn a mapping from a high-dimensional continuous input space to a low-dimensional, discrete output space, a clear example of a dimensionality reduction method. In other words, it automatically creates a set of organized categories based on the data it observes.

Many artists who claim to make use of neural networks are in fact using SOMs. Such is the case of sculptor Yves Amu Klein, who demonstrates an explicit commitment to creating autonomous robotic life forms. His *Living sculpture* project, a series of works that attempt to “bring emotional intelligence and awareness to sculptured life forms” (Klein 1998, 393), directly resonate with Burnham’s vision.

Many of Klein’s works show adaptive features. Such is the case of *Octofungi*, a 1996 robotic sculpture that relies on shape-memory alloy wire to control eight robotic legs arranged in a circle. The movements of the robot are defined by the interaction between the position of the legs and the value of eight photocells that measure incoming light from all directions. The data from both legs and photocells is fed into a SOM which autonomously extracts regularities from the input data

[36] Kohonen, which I briefly introduced in section 3.1.6, is a good example of a scientist whose research in connectionist adaptive systems, shelved during the AI Winter, would finally be recognized during the 1980s.

space and chooses to activate one of the eight legs as a response. The SOM thus learns from its environment, adapting to it in real-time.

Other works from Klein's *Living Sculpture* series which make use of SOMs, include *Scorpibot* (which was the artist's first attempt at using them), *The Pods*, *Bella*, and *Flexicoatl*. Klein loosely refers to them as "neural nets" which is somehow misleading since, as I pointed out, the expression does not refer to a specific technique but is rather an umbrella term for a large group of artificial models that try to achieve adaptive properties by reproducing biological neural networks on computer systems.

Nicolas Baginsky also uses the term "neural nets" interchangeably for SOMs when describing his piece *The Three Sirens* (1992—2005), a robotic music band who play improvisational rock music. The band consists of three robots: a slide guitar player, a bassist and a drummer. The guitarist and the bassist use SOMs to direct their actions, playing live music in response to the sound environment they generate in real-time. Since the sound environment is largely influenced by their own playing, they are also entangled in a feedback loop that runs through their bodies and their environment. A particularly fascinating aspect of the piece is how the robots have allegedly evolved through a number of years as the connections of their SOMs were preserved between performances:

> When the robot first started playing in december 1992, the six neural network [sic] that control the machine's behaviour were randomly initialised. Today there are several different sets of networks available for different modes of operation (different speeds and tunings). All these sets are descendants of the primal neural nets from 1992. This means that the robot system now has the experience of about ten years of playing. Not constantly but regularly. (Baginsky 2005)

Quite interestingly, these systems use the self-organizing properties of SOMs as part of a decision-making process, which is in counterpoint to what these models were originally designed for. Their ability to organically remap their inputs into outputs in a meaningful way seems to be effectively used by these artists to generate novel behaviors that is both organized, yet definitely nonhuman. Whereas this particular approach of using unsupervised learning to control an agent-based system is an original, creative hijacking of the technology, the field of ML has developed a distinctive approach for training agents.

Reinforcement Learning

Reinforcement learning (RL) (c.f., section 3.2.1) tries to address problems involving an agent that attempts to take actions in an environment in order to maximize its reward over time (Sutton and Barto 1998). The agent learns by taking actions and receiving positive or negative feedback from the world through rewards. A reward is a single-value information unit given to the agent in response to his state or actions. Following Holland's definition of adaptation, the goal of a reinforcement learning agent is to modify its inner structure in order to maximize its performance — represented as the rewards it collects over time — as it evolves in the environment (Holland 1992).

The field of Reinforcement Learning (RL) emerged in the late 1980s as the result of a coalescence between behavioral psychology, optimal control theory and dynamic programming. In reinforcement learning, agents evolve inside an environment defined as a discrete time-based stochastic control procedure known as a Markov decision process (MDP). In this procedure, an agent takes actions in the environment based on what it observes. Each action modifies the environment, yielding a new set of observations for the agent as well as a single-valued reward feedback. The goal of the agent is to maximize its rewards over time. In order to do so, it usually proceeds by trial-and-error, trying to infer what is the best course of action to take in a given context based on rewards and punishments received in the past.

An example of a Reinforcement Learning technique is Q-Learning (Sutton and Barto 1998, 148), a procedure in which the agent bases its decisions on an estimator function called a Q-function. This function takes as parameters both the observation s and an action a and produces an estimate of the expected reward the agent will get for taking action a in context s:

$Q(s, a) =$ expected reward for taking action a given observations s

After each action taken, the $Q(s, a)$ function is slightly adapted by the agent to give a better approximation of the expected reward in the future. There are a certain number of ways the agent can use that information to choose the actual action it is going to take. The way the agent uses the Q-function to choose its actions is called a policy. The most obvious policy is just to take the action with the maximum Q-value:

$a = argmax_{a'} Q(s, a')$

This is what we call a greedy policy. However, remember that the $Q(s, a)$ function is learned and thus, it is not the actual expected reward, but rather, an approximation of it based on what the agent has been observing in the past (i.e., in the state-action pairs it went through). A purely greedy policy favors exploitation of what the agent already knows, which is done at the expense of exploration. Concretely, greedy agents will tend to get stuck in the same, safe zone where they started, because they have not been given a chance to try out different things (i.e., to wander inside the whole state-action domain).

To alleviate this problem, which will more often that not result in sub-optimal behaviors, one has to introduce some exploration in the policy. A simple way to do this is to let the agent be greedy most of the time but, once in a while — say, with a probability of ϵ — let it take a random action. This variation on the greedy policy is called the ϵ-greedy policy and is the one used most frequently in reinforcement learning. This is due both to the simplicity of its implementation as well as to its surprising efficiency in allowing the agent to converge to a good solution in most situations.

In their 2014 installation *Zwischenräume*, artist Petra Gemeinboeck and computer scientist Rob Saunders looked at the live adaptative performances of robotic agents. The robots are “sandwiched” between the gallery wall and a temporary wall. Each one of them is equipped with a motorized system that allows it to move vertically and horizontally, covering a specific region of the wall. The robots are also equipped with a puncturing device that allows them to make holes through the surface, as well as a camera and a microphone that allows them to sense their environment. The system also give robots the ability to extract features from the camera and from the audio signal. It combines all this information using both a Self-Organizing Map to detect similarities between images, and a Reinforcement Learning program that tries to “maximise an internally generated reward for capturing ‘interesting’ images and to develop a policy for generating rewards through action”. The level of interest in the described system is based on a measure of “novelty and surprise” where “‘novelty’ is defined as a difference between an image and all previous images taken by the robot” and “‘surprise’ is defined as the unexpectedness of an image within a known

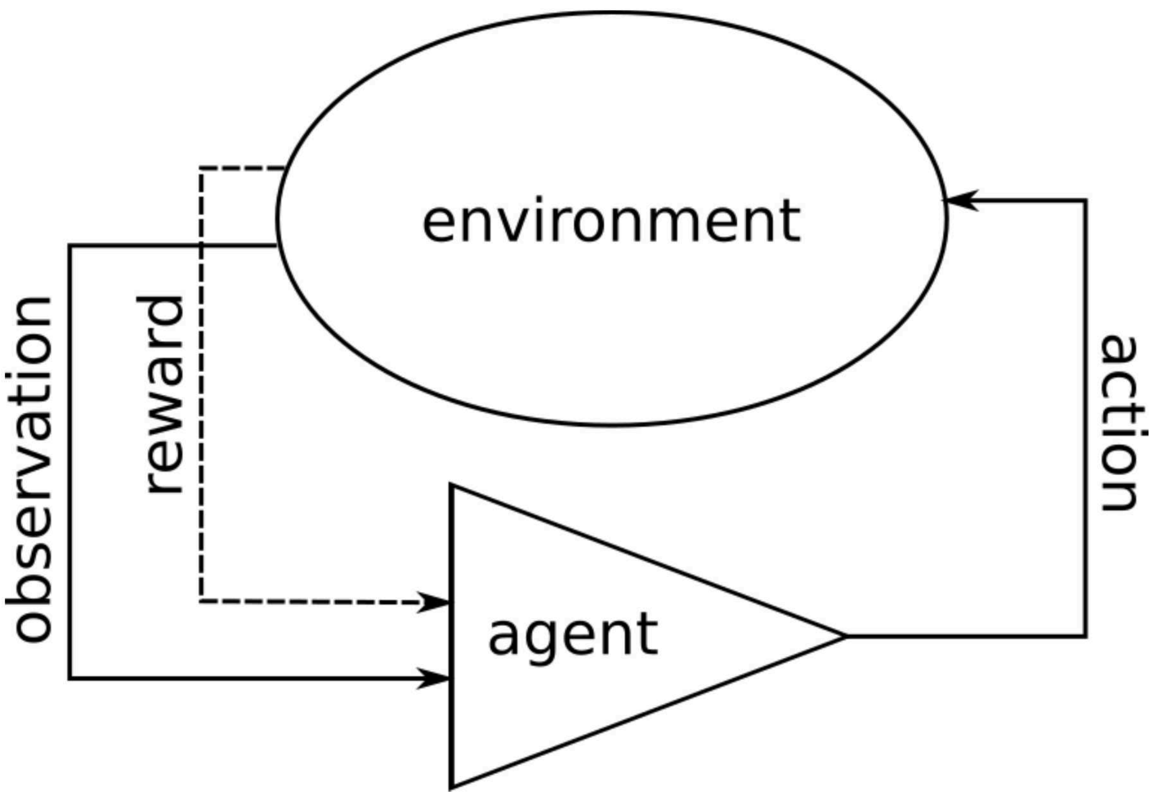


Figure 17: Schematics of the Reinforcement Learning feedback loop. At each step of the process, the agent takes an action in its environment. In response to this action, the environment returns a set of observations as well as a reward measuring the performance of the agent. The goal of the agent is to find an optimal policy (i.e., a decision mechanism to assign appropriate actions in response to observations) that allows it to maximize its rewards over time.

situation" (Gemeinboeck and Saunders 2013, 217).

An interesting aspect of the work is the relationship that is established between robots and audience. As they dig holes through the walls, the curiosity-motivated agents become an "audience to the audiences [sic] performance". It is thus "not only the robots that perform, but also the audience that provokes, entertains and rewards the machines' curiosity" as the "robots don't only respond or adapt to the audience's presence and behaviours, but also have the capacity to perceive the audience with a curious disposition." (218)

Canadian artist, musician and AI researcher Stephen Kelly, one of the collaborators on *Vessels*, has produced a number of experimental works using Genetic Algorithms and Genetic Programming (GP). Genetic Programming is a particular approach to GAs where the individuals that get evolved are instances of computer programs. In a typical GP application, populations of such programs are generated, tested on a problem, and then selected based on their performance. The fittest candidates are used to generate new offsprings using different genetic manipulations. Hence, Genetic Programming is considered a form of *policy search*, where the agents' behaviors are evolved directly based on their performance over a given task — as opposed to *value search* methods such as Q-learning where agents are rather made more efficient based on the adaptation of a value function that tries to estimate what is the best action to take in a given context (Grefenstette, Moriarty, and Schultz 2011).

Kelly's *Open Ended Ensemble* is an ongoing project involving physical, sound-generating agents. In the current version (labeled *Competitive Coevolution*), two robotic probes move along a fluorescent light fixture, trying to find the region with the lowest electro-magnetic radiations.

> The agent's behaviour is adaptive, subject to an evolutionary process in which a random population of computer programs slowly evolve, eventually achieving enough control of the robotic probe to coax its movement away from the source of radiation and into silence. Meanwhile, the light fixture would prefer to maintain the drone, and slowly evolves a strategy of its own, learning to move the lights and trap the probe in regions of strong radiation. An arms race ensues as the two competing forces interact and coevolve, akin to predator/prey or host/parasite relationships in biological systems. (Kelly 2016)

Kelly's strategy, in this particular version, echoes my own approach for staging agents with

Figure 18: *Open Ended Ensembles (Competitive Coevolution)* (2016) by Stephen Kelly. Hamilton Artist Inc, Hamilton, Canada. Image courtesy of the artist. Photo by Caitlin Sutherland.

conflicting goals, such as in *Drift* (2007) (solitude vs company) and *Fifth Absence* (2011) (energy need vs desire of shade). The agents in *Open Ended Ensemble* have an imperfect control over their movements, and their observations are limited, which places them into a partially unpredictable environment.

The artist reported that one of the biggest challenges in creating this work was the fact that the plastic, visual and audio components of the piece were, in his opinion, taking much more place in the aesthetic space of the piece than its behavior, obscuring the trial-and-error process. This remark resonates with my own observations working with adaptive agent-based systems in art. It is not clear at all how a learning behavior can be observed or felt by the audience while integrating it as part of an experience that manifests through different media in the creation of a global experience. I have experienced similar difficulties working on projects using RL such as my underwater installation *Plasmosis* (2013) and *N-Polytope: Behaviors in Light and Sound After Iannis Xenakis*, which is described in more detail in chapter 6.

3.2.2 Components of a Machine Learning Algorithm

Parallel to the category of task they are designed for, Machine Learning algorithms can be qualified by the interoperability of four constituents: (1) the *model*; (2) the *optimization procedure*; (3) the *evaluation function*; and (4) the *data*. The optimization process gradually improves the model based on its performance over the data using an evaluation function (Alpaydin 2004, 35–36). This is roughly true for each kind of task, with many variations within the kinds of techniques that are suitable for each of these components.

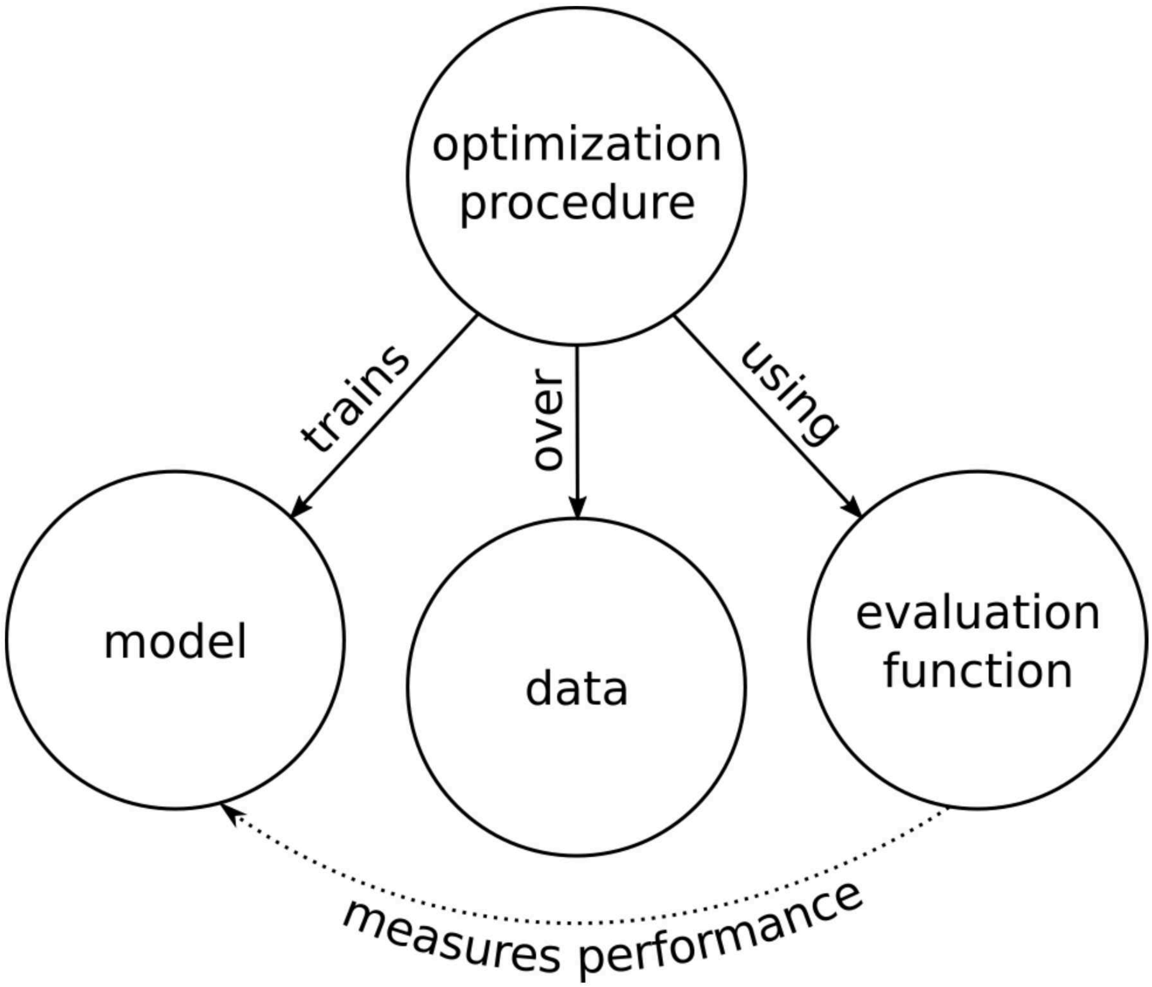


Figure 19: Relationship between components of a Machine Learning algorithm. The optimization procedure trains a model over a set of data/examples using an evaluation function that measures the performance of the model.

Models

Models in Machine Learning refer to the computational structure that gets modified through learning. The best way to think of a model is as a function that tries to approximate as close as possible a distribution of data, based on a sample of that distribution (the dataset). The model contains free parameters that are to be adjusted by the training algorithm. For example, in the Multi-Layer Perceptron, the parameters are the "weights" or "synapses" that connect the neurons with one another. Other models include decision trees, Bayesian networks, Support Vector Machines and nearest neighbors models. In a GA, the model is the function that associates DNA strings with a phenotype, while the chromosomes are the free parameters to adjust.

Models are the object of important debates in the field of Machine Learning, being the defining flagships of different research strands. However, when it comes to artistic works, they are possibly the least explored dimension. As was expressed earlier, most adaptive artworks involve a very restricted set of models, which happen to be among the most easily understandable and applicable ones (GAs and SOMs). This most likely has to do with the fact that scientists and artists have different goals and expectations. To put it simply, an apparently small improvement in the performance of a model can be seen as revolutionary from a scientist's perspective but will not change much in terms of how it affects the experience of an artwork.

Nonetheless, there are at least three ways in which models can affect artistic outcomes. First, the nature of the model is often an important part of the concept of a piece: the imaginary space opened-up through the use of neural nets differs conceptually from that of evolutionary computation or decision trees. For example, Sims' *Galápagos* plays with the richly evocative nature of evolution, allowing the user to take part in a story of genetic adaptation as the godlike subject that runs the natural selection process. Ben Bogart's installation *Dreaming Machine #2* (2009) and Ralf Baecker's *Mirage* (2014) both involve neural networks in pieces about memory and dreaming — two themes that lie at the center of research on neurology that directly inspired computer-based connectionism.

Secondly, models have specific structures that allow different forms of "hijacking". Chapter 4 presents an artistic strategy whereby a Genetic Algorithm model is used in a way utterly different than what it was designed for. Google's *DeepDream* is a good example of a creative approach that employs specific properties of a neural net to transform it into a generative device it was never meant to be. These artistic strategies usually take advantage of an accidental feature of a model, diverting it out of its habitual or intended use. It requires a good comprehension of the model and/or an experimental approach.

The third process by which models can impact artistic works is more subtle and has not been the object of serious analysis. It has to do with the fact that different models will yield, or afford, different kinds of behaviors. The variety and types of behavioral strategies that the model allows, and the "smoothness" — or "abruptness" — in the evolution of these strategies during learning, are examples of how models can affect agent aesthetics.[37]

Optimization Procedure

The optimization procedure — also called *search* or *training algorithm* depending on the context — changes the parameters of the model in an attempt to improve its responses over time. Different kinds of such procedures exist, each with their own advantages and domain of application. For example, there is a vast amount of research on training algorithms for neural networks, using different optimization approaches such as Stochastic Gradient Descent, Genetic Algorithm, and simulated annealing.

Most optimization algorithms exploit the Cybernetics notion of negative feedback: in response to the perceived error yielded by its actions, the organism adjusts its inner structure in a timewise manner, step by step, moving towards an optimum. Whereas many cyberneticians were interested in the process itself, for scientists working in the field of Machine Learning, optimization is a means to an end. The principal goal is to train a system that will perform well on a particular task *once it has been optimized.* What happens before that, the behavior of the system as it gets there, is considered irrelevant. Conversely, it is what is probably the *most* relevant to an aesthetics of adaptive behavior.

[37] The advantages and disadvantages of neural nets, as opposed to GAs and other techniques such as fuzzy logic and Support Vector Machines, is a broadly debated topic in the field of Machine Learning.

The learning process can typically be fine-tuned using a set of meta-parameters. For example, most optimization methods involve the use of a *learning rate* parameter which represents the speed at which the system moves towards a local minimum. There is, however, a trade-off: high learning rates will get models to converge faster, however, they will often yield poorer results; lower values will take more time but result in a finer model. A common way to solve this trade-off is to start with a larger learning rate and slowly decrease it over time.

Another example is the *exploration vs exploitation* dilemma in Reinforcement Learning (Sutton and Barto 1998, 4). When an agent moves in a space searching for the best strategy to maximize its reward over time, it needs to be able to both exploit its current knowledge (by making decisions it thinks are going to yield good rewards) and explore new avenues (so as to avoid getting stuck in a region of the space that yields poor rewards). Exploration is usually more chaotic and random, while exploitation is targeted and greedy. In a typical RL setup, agents will start by exploring and, over time, be tuned to favor exploitation as they become more efficient in accumulating rewards.

The agent's tendency to favor exploration over exploitation is usually represented by a single parameter. For example, in one of the most commonly used learning policies, called ϵ-greedy, a parameter ϵ between 0 and 1 represents the probability that, at any given step, the agent will choose a completely random action (if not, then it will choose the action it believes will yield the highest return, hence the name "greedy") (28). Altering ϵ can be used as an aesthetic trick in agent-based systems, allowing the shaping of behaviors in real-time, moving them between chaos and order. This strategy was applied in the immersive installation/performance piece *N-Polytope* (2012), for the construction of live generative behavioral patterns, as described in section 6.3.4.

Evaluation Function

The *evaluation function* measures the performance of the model in completing its task. In Supervised and Unsupervised learning, it is usually referred to as the *loss function* or *cost function.* In a classification task, for example, the category predicted by the model given an example to classify is compared to the expected target category: the more the model misses the target, the larger the loss. In Reinforcement Learning, the evaluation function is called the *reward function*, while in

Genetic Algorithms, it corresponds to the *fitness function.*

Among the three dimensions of a Machine Learning algorithm, the evaluation function is probably the one that is the most readily useable by authors. This is because it has been designed specifically for the purpose of bringing human input into the equation. Models and optimization procedures are meant to be rather agnostic: the evaluation function determines the kind of "problem" one tries to solve. However, the approach in art completely differs from that of science. While scientists use evaluation functions as objective criteria for the learning algorithm to solve, artists typically use the evaluation function as a tool for generating self-organizing behaviors, subject to their own authorial control. In other words, for scientists, the evaluation function represents the goal they aim to achieve, without any care for the way it is reached (i.e., the goal is more important than the process to reach it), whereas for artists the relationship between the evaluation function and the goal (which is to generate interesting behaviors) is indirect (i.e., the process *is* the goal).

Artists can thus play with evaluation functions and observe how the agent responds. An evaluation function can also be learned or attributed by another agent. Finally, evaluation functions can be interactive, with either the artist or the audience replacing the function by directly giving an evaluation of the system's performance. In the case of evolutionary computation, this technique is known as an Interactive Genetic Algorithm (IGA), an approach first proposed by Richard Dawkins (Dawkins 1986).

Karl Sims' *Galápagos* (1997), which was presented earlier in section 3.2.1, is one of the most renowned examples of the use of IGA in an interactive installation. Here, visitors are asked to select their favorite artificial 3D creatures, whose genetic code is used to create the next generation through mutations and crossovers. Core to the work's aesthetics is its participatory nature, engaging audiences in the production of novel forms through a playful and intriguing experience.

The *Fifth Absence* (2011) is another example of how an evaluation function can be used poetically in the generation of an artificial behavior. As described earlier, the work involves a robotic agent immersed in a behavioral conundrum through the implementation of a reward function precisely designed to generate it. The agent in this artwork is forced to discover, through trial and error, a strategy that will allow it to match its desire to avoid looking at light sources with its need to get solar energy. The slow-paced behavior of the agent, who moves about once every 2–3 minutes, places it in a different category than *Galápagos* in terms of aesthetics. Like most other interventions in *Absences*, this is a very conceptual piece, as the shape of its behavior can not be perceived in real-time by human subjects and thus needs to be imagined by the audience.

Data

Data is an often overlooked, yet crucial dimension to consider when thinking about adaptive behaviors, especially in an artistic context. There are practical concerns when dealing with data encoding, as well as challenging issues that arise when dealing with high dimensional spaces, such as is the case with image or speech recognition, which are largely beyond the scope of this dissertation.

The first thing to consider in regards to data is the kinds of inputs and outputs that will be fed into the system — in other words — what the agent will be able to observe, and how it will be able to respond to these observations. In order to be effective, these inputs and outputs need to be carefully chosen to afford the kind of experience the artist has in mind. Moreover, there needs to be a way for the agent to make inferences, otherwise no learning will happen. For example, a system that can only detect light cannot be asked to learn about the sounds made by visitors.

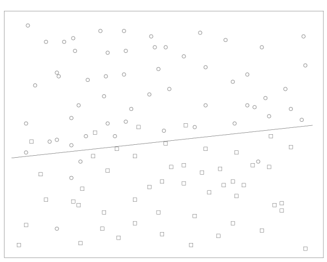
The set of sensors/observations/inputs and actuators/actions/outputs, and the way they are embodied in the adaptive physical devices that are staged in an agent-based artwork, possibly constitute the most important decision an artist has to make in the creative process, as it will define the kind of space in which the agent can evolve, the sort of behaviors it can afford.

Secondly, it is self-evident that the data distribution from which the examples are selected has an important influence on the reactions and establishment of the system's behavior. One of the most dreaded issues in Machine Learning is *overfitting*, a problem that arises when a system estimates "too perfectly" a specific dataset, thus becoming less efficient at making predictions on unseen samples (i.e., taken outside of the training dataset). While overfitting is a plague for data scientists, it might actually be exploited creatively by artists, by hand-picking data (such as by creating a constrained environment) in order to encourage a specific response in the system.

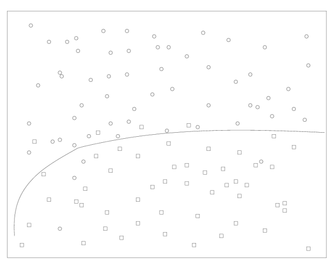
Other Considerations

Both from a scientific as well as an engineering perspective, Machine Learning techniques are simple in spirit, yet extremely complex when it comes to details. Many elements can influence the success or failure of a particular algorithm on a particular problem, and much energy is spent in the field to compare strategies and try to extract general principles behind learning.

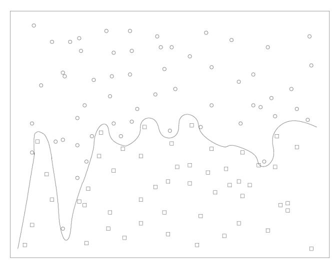
The biggest challenges are related to issues that arise when dealing with high dimensional data, which becomes the case when dealing with image or speech recognition. These difficulties mainly concern questions of generalization, that is, the problem of training a model on a specific set of examples so that it becomes good at making predictions when faced with examples taken outside of that dataset. An important conceptual issue is known as the *curse of dimensionality.* It spans many unique problems that arise when dealing with high-dimensional data. One of the most fundamental consequences of the "curse" is that the number of free parameters (which amount to the representational power of the model) need to be tuned according to both the dimensions of the input space and the size of the training database.[38]



(a) Linear model (e.g., perceptron).



(b) Nonlinear model (e.g., MLP).



(c) Nonlinear model overfitting the data.

Figure 20: Example comparisons of how different kinds of model classify data points in a two dimensional space, including a case of overfitting.

It is largely beyond the scope of this dissertation to give a detailed account of these theoretical concepts. However, artists should be aware that these techniques require at least some basic knowledge if one wants to be able to manipulate them as creative tools and material. Unfortunately, there

[38]The curse of dimensionality is directly linked with issues of overfitting and its opposite, underfitting.

exist almost no resources at the moment specifically dedicated to teach artists about ML,[39] and most of the tutorials require at least some degree of knowledge in mathematics and programming.

3.3 Conclusion

The history of computational synthetic agents after the post-WWII era runs across several disciplinary grounds. In this chapter, I examined this history through consideration of the notions of adaptation and learning, mainly from the point of view of media art history and computer science.

A few features of these histories stand out, and need to be highlighted. Tensions between opposing ways of thinking about body, mind, life, and intelligence act as the backdrop for these historical markers. Computationalism — the concept that cognition is the same as computation, that software precedes hardware, and that the Turing test is decisive in determining if an agent is cognizing or not — is central, and has been espoused in particular by symbolic, "good old fashioned" AI. Opposed to this computational theory of mind are views that argue for the importance of the body in the performance of a cognizing system. Somewhere in the middle, are connectionist approaches, which claim that intelligence is all about learning subsymbolic, statistical relationships between the agent and its environment.

Cybernetics had a significant influence on the evolution of contemporary art in the 1960s. Thinkers such as Roy Ascott and Jack Burnham explained the transformation of art in these years as a displacement of the aesthetic locus from objects to processes, and described the art world itself as a flow of information and behaviors between a multitude of systems.

Adaptive devices were central to these visions. However, as the sweeping influence of rule-based AI from the 1960s to the mid–1980s pushed away alternative approaches, it seems that the importance of adaptive and learning systems in contemporary media art was equally diminished.[40] Still, there remains a noticeable strand of artworks based on learning agents that runs through history, dealing with similar questions and facing similar challenges.

[39] One promising initiative is a book in preparation by computer scientist and artist Gene Kogan, in collaboration with designer and artist Francis Tseng (Kogan and Tseng 2016).

[40] As we will see in chapter 5 , in the 1980s–1990s, artists seemed to have moved away from AI towards the field of Artificial Life.

In the second part of the chapter, I examined Machine Learning algorithms as described in scientific literature, in an effort to reveal their fundamental mechanisms, with the objective of bringing out their aesthetic potential. I first gave an overview of the three different categories of tasks that can be addressed by ML algorithms: *supervised*, *unsupervised*, and *reinforcement* learning. Secondly, I described the different components of such systems and how they interact. In a typical ML algorithm, an *optimization procedure* is used to train a *model* over *data* using an *evaluation function* (Alpaydin 2004). I provided several examples on how these different elements can be, and have been, exploited by artists working with adaptive agents.

These characteristics are not only important for understanding how such methods might have aesthetic repercussions, but also demonstrate how difficult it is to seamlessly move between the lower level of choice of algorithm (or technique) and the higher phenomenological level of aesthetic experience — something that I will pick up in the next chapter.

A crucial consideration is that ML algorithms are designed for optimization, which makes their use in art counter-intuitive. Yet, there are multiple ways in which artists can appropriate these techniques by doing what they do best: diverting the technology from its targeted application. There is no such thing as the "best" or the "most aesthetic" behavior for a computational agent, therefore there exists no objective evaluation for it. Yet, for example, it is possible for artists to "toy with" the evaluation function of a learning system as a way to generate effects, or to use the fundamental properties of a model as a conceptual tool, or to choose the training dataset carefully to produce specific content.

This set of tools allows one to better understand agent-based works such as the *Absences* series. However, one of the characteristics of these works is that they were designed for the outdoors, with no intention of being shown in front of an audience, aside from their documentation. What happens when an adaptive agent-based installation is presented in front of a human audience? What effects are generated? How does the public react to such behavioral patterns?

The next chapter describes a collaborative robotic work that was created between 2010 and 2015. Called *Vessels*, it involves a fleet of small autonomous water-dwelling agents, which create an emergent, social behavior through their adaptation to their milieu and their peers. The piece, which can be shown either indoors or outdoors, allows the audience to directly observe their evolution as they respond iteratively to their ecosystem. The question thus remains: what new kinds of aesthetic paradigms do adaptive systems produce or generate in human perceivers, who themselves are inexperienced in regards to the behaviors of such nonhuman agents and processes?

Chapter 4

Vessels

> Empty your mind, be formless, shapeless - like water. Now you put water into a cup, it becomes the cup, you put water into a bottle, it becomes the bottle, you put it in a teapot, it becomes the teapot. Now water can flow or it can crash. Be water, my friend.
>
> – Bruce Lee

November 13, 2015. The sun is setting upon the solemn campus. In front of the Law School, the Goddess Athena opens her arms to the “Nouvelle Cité”, heritage of the 1960s French marxist revolution. Paris was attacked yesterday by groups of armed men. But we are in Strasbourg, and the turmoils of last night have already settled, though a spectre of this violence seems to be walking among us.

A woman asks if the robots are there in memory of the victims. She explains that she lives in one of the condo towers in front of the fountain and that she noticed, from up there, the slow oscillation of colored lights on the water. People have gathered on the eastern side of the basin, where the wind was gently pushing the bots. A small grouping of the robots is slowly flickering in greenish blues; over there, three orange ones are hopping, moving in short, erratic bursts, their motors roaring like a voice.

There, on the other side of the fountain, a single purple one looks like it’s slumbering, it’s pale color slowly undulating. Parents are watching their two kids playing around it, gently pushing it when it reaches the border.

Insect sounds, responding to one another, like a choir of electronic crickets; whispers coming from the audience, as new groups of spectators gather around the place.

A gust of wind blows, pushing them away. By now, menacing clouds have also appeared: pink and orange, tinted by the smog-drenched light of the setting sun. Lighting bolts flash over the horizon: a storm is coming.

Fading colors.

The purple one has reached the rest of the group. All of a sudden, it starts squeaking now like a fax machine. Three of them respond, then more. Binary codes echoing through space, a ballet of colorful beings, all converging together as the first drops of rain fall from the sky. Slowly, their color changes, a new community is formed as the territory is collectively redefined. People start leaving, kids want to stay but they are scolded as the air gets thicker.

They take a break now. This one has started shining in a strange, whimsical way. This other one shakes and pushes the other two. Their colors are now moving between green and orange. Suddenly, the whole group starts moving towards the shore in unison. They bounce against the border, pushing, as if in some kind of a panic.

Then, as quick as they began, they all stop. One of them leaves, changing back to a blueish color. Then another one. And another. A sudden squeak, followed by screams of noisy binaries: they all split up.

The clouds have passed. But everybody’s gone.

Figure 21: *Vessels* (2015), L'Ososphère, Strasbourg, France. Photo: Philippe Groslier.

This chapter examines the research and creation process behind the realization of *Vessels*, an artistic robotic installation consisting of large collectives of water-dwelling mobile robots, created as a collaboration between myself, Samuel St-Aubin and Stephen Kelly from 2010 onward. As they move over the water's surface, the bots engage in different forms of social interplay, influencing each other's behavior and appearance in oscillating movements of convergence and divergence. Moreover, each robot perceives a specific dimension of its environment such as water quality (air temperature, atmospheric pressure, or ambient light and sound) which influences its behavioristic character. For example, a high temperature measured by one of the agents could make it increase its speed or give it a preference for rotating clockwise. In turn, this individual change in behavior contaminates its neighbors' demeanor.

Over time, a collective behavior that is specific to the immediate environmental characteristics of the presentation site emerges from the agents' socialization. The work thus acts as an organic laboratory that responds to hidden features of the urban ecosystem by displaying emergent social behaviors, offering the viewers a new perspective on their living milieu and a model for cultural exploration.

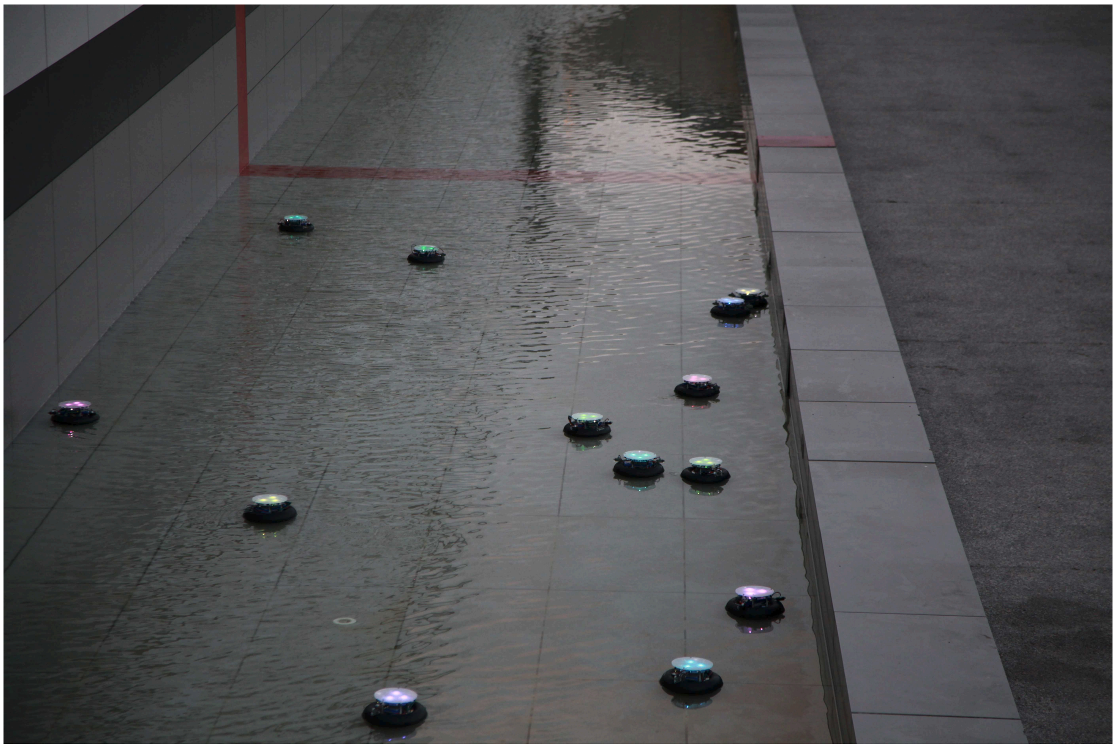


Figure 22: *Vessels* presentation at LABoral Centro de Arte y Creación Industrial (Gijón, Spain). August 2013. Photo by Sofian Audry.

Here, I focus on the computational aspects of the project, in particular the use of Machine Learning techniques. *Vessels* provides a valuable case study for applying the concepts developed in the previous chapter. Specially, it highlights the different challenges faced when making use of such techniques in behavior-based artistic works. Furthermore, the project offers an example of the kind of aesthetic effects that can be generated through such work.

One of my intentions in creating this work had been to use Reinforcement Learning as a way to generate different kinds of behavior in the robots. I explain how and why this approach was eventually abandoned for both technical and artistic reasons. In fact, it has already been pointed out that the traditional inclination of AI towards problem-solving and optimization makes it unpractical for creative applications (Eigenfeldt, Burnett, and Pasquier 2012; Mateas 2001). RL is particularly challenging for artistic works because it is caught in a stark paradox: the traditional contexts in which media art installations are produced and presented are particularly ill-suited for Machine Learning in general, and Reinforcement Learning in particular. This is mainly due to the fact that RL agents need to be exposed to a lot of data in order for them to learn. However, in a gallery setting, the audience typically gives its attention to works of art for only a few minutes, which is usually not enough for the learning process to complete.

I follow up by describing how we successfully made use of genetic algorithms by hooking into the learning step as a way to evolve cohesive forms of behavior in real-time. GAs are the most-used strand of ML methods employed in the creation of media art.[1] In *Vessels*, we use GAs in a very specific manner: as a way for robots to collectively self-organize so that a form of behavioral "family resemblance" emerges from their interaction with one another and their environment.

Following these observations, I claim that one way of using Machine Learning methods within the context of agent-based artworks is to provide an algorithmic framework which allows for the generation of adaptive *behaviors* rather than the production of an efficient solution to a definite problem. Most learning algorithms define an iterative procedure where a model is refined at each step towards achieving a certain goal. By hooking into this process, an artwork can reveal to the audience the process of adaptation itself, which can be made to be artistically compelling through its evocation of familiar behavioral patterns usually displayed by living and/or sentient beings.

The first section of the chapter describes the broad artistic goals of the project. I address issues related to form, content and choreographic development in a decentralized collective of adaptive agents. The second section provides an overview of the technical dimension of the work, with some contextual insights on the practical reasons justifying the choices that were made. The third section depicts the research-creation process, bringing in questions of methodology as well as both practical and theoretical implications of the work. The chapter ends with a description of similar

[1] For examples of ways GAs have been used in visual arts and music, see (Johnson and Cardalda 2002).

artistic installations involving collectives of agents and analyses the affective experiences of *Vessels* as reported by the audience.

4.1 Artistic Intent

Interactive media artist and theorist Simon Penny stresses the importance of establishing a relationship between the work, the viewer and the environment they share in works of art based on artificial agents (Penny 2000). In *Vessels*, this interrelation is constructed from behaviors that occur not only at the individual level, but foremost at the global, emergent level. The intention here is to develop a kind of adaptive and distributed choreography; a social dance with no central conductor that induces and maintains viewers' interest by evolving constantly. My intuition is that this will bring a sense of aliveness that will allow the audience to more intimately relate to the work, identify with it, and ultimately to inspire the audience to reflect and question their own relationship with their living environment and its inhabitants.

We addressed these challenges through three (3) major artistic objectives: (1) to dynamically occupy space; (2) to create social interactions between the robots; (3) to establish a relationship with the environment. These goals were deliberately meant to be blurry, allowing room for experimentation and creation. They are also incomplete and cannot, in themselves, fully encompass the conceptual scope of the project. They serve primarily as anchor points and reflection axes for supporting the creative process.

4.1.1 Occupying Space

Our general vision of the spatial disposition of robots was to create a constantly evolving mixture of densities and colors, suggesting different configurations of both collective and individual behaviors. Agents in *Vessels* operate in territories over which spectators have a global perspective. One of the important dimensions of the work is thus the way by which the robots spread across space. The robots must neither constitute a cloud of detached individuals that moves randomly, nor must they be assembled in a single, static mass. Their locations and movements should be diverse such that at any given time, the surface on which they dwell might display contrastive variations in density and movements that evolve over time.

The idea of varying the ways robots occupy their living milieu is tied to one of the important challenges described by Penny, which is to ensure that an agent's behavior is differentiable from pure randomness while still evoking a sense of mystery and uncertainty. Discussing his experience in creating *Petit Mal*, a piece featuring a robot interacting with users, he explains that a fundamental concern for artists is to generate "poetic richness which is clear enough to orient the user but unclear enough to allow the generation of mystery and inquisitiveness" (Penny 2000, 441).

These considerations illustrate one of the biggest challenges in this project. The establishment of complex rules in the system contribute to the unfathomable character of the work, but may also come to unbalance the emergent behavior by making it indistinguishable from the purely random.

4.1.2 Generating a Social World

While the spatial distribution of robots in *Vessels* is related with sociability as the natural outcome of individual inclinations, these dynamics are assuredly not the only aspects of the robots' social life that can be explored as evocative material. We decided early on that agents should have a way to communicate with one another to generate different kinds of social actions and events. Following the same idea of a bottom-up approach, we designed a very simple process involving "social acts"; kind of atomic behavioral building blocks by which robots could interact with one another.

Social acts would be triggered by a single robot who would lead the action through three (3) different phases: negotiation, action and release. At first, the leading agent would call for peers to perform the action. Those receiving the call would not send any messages back but would either ignore it or start acting as temporary followers of the leading bot. While engaged in this follower-followed relationship, the leading bot would have the option, from time to time, to activate a social event. Examples include following or evading the leader, spinning clockwise or counterclockwise, blinking lights or emitting sound. Finally, the leading robot could call off the action, releasing its followers to their own inclinations.

4.1.3 Interlacing Identity and Environment

Social interactions between the agents evoke a sense of a community, allowing the audience to identify with the work. A final objective of this work is to build on that bond, to offer a space for the public to reflect upon the urban environment by rooting the agents' identity and behavior in the human audience's perception of underrecognized characteristics within their own milieu.

Installing the robots in an aquatic environments is directly linked to this ambition. Water carries a strong poetic evocation, being both the quintessential source of life on earth, and sometimes providing an unfathomable, possibly dangerous territory. Ponds and harbours in the urban environment are places of gathering and reflection: places where one can evade the turmoils of city life to meditate, observing birds and insects moving on the surface, listening to the sounds of wind and lapping, dragging oneself into the ripples and vortices deep below the surface. Furthermore, the water element contributes to an impression of fragility in the robots and makes their movements imprecise, which gives them more personality.

Robots' reaction to their environment is guided by a simple aesthetic principle: behavior displayed by the community should reflect the environmental characteristics of the location, thus acting as kind of a dynamic "signature" of the milieu. This general rule can be expressed more or less as the result of the tension between two sub-principles, namely; (1) an *identity principle* stating that a community of robots presented in a given location, showing a specific set of environmental features, will display an emergent behavior *identical or similar* if it is presented again in the same conditions; and (2) a corresponding but opposite *contrastive principle* pushing robots towards engaging in a *different* behavior when subjected to noticeably distinct environmental conditions.

In line with our bottom-up methodology, the relationship between the agents and their environment is first and foremost expressed at the individual level. As previously mentioned, each robot is equipped with a sensor that measures specific environmental data, such as air temperature, atmospheric pressure, carbone dioxyde level, ambient light intensity or audio noise level. That specific piece of data directly influences the personality of the robot, who adapts to it by transforming itself each time a new measurement is performed.

If we were to just leave things that way, then the individual robots would each come to modify their behavior according to their own sensor; upon convergence, a robot would have the "personality" of its sensor, so to speak. However, what we really want is that the community *as a whole*, rather than as distinct individuals, tends towards behaviors that render the dynamic state of the ecosystem it inhabits. The personality of an agent at any given moment thus needs to be tainted not only by their own environmental measurements but also by that of other agents. As I explain later on, this will be made possible through adaptive strategies. The relationship between the robotic society and its environment is thus assembled through the constant negotiation between individual and collective identities.

The cohesive inhabitance of a space, complex socialization, and an embeddedness in the environment; these three encompassing goals have provided a supportive frame for the development of *Vessels*. In the next section, I lay the groundwork for an in-depth examination of the research-creation process by broadly describing the technical components of the work in relationship with these goals.

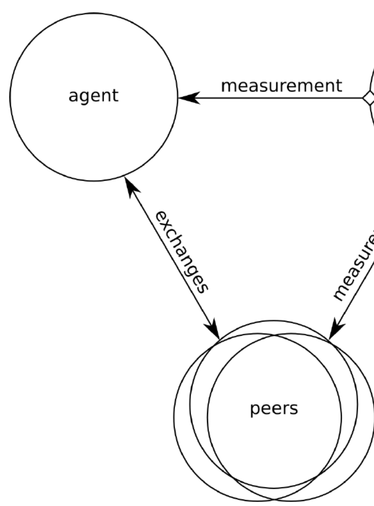


Figure 23: Schematic representation of the relationship between robotic agents among each other and their environment. Agents collect different measurements from their environment and exchange this information organically through socialization.

4.2 Technical Overview

The artistic intentions outlined in the previous section are intertwined with practical considerations about the physical, hardware and software dimensions of the robots. Though this research focuses mainly on the algorithmic aspects, I want here to give a sense of the agents' capabilities by taking a peek at their physical constituents.

Vessels' robots were designed iteratively over the course of several years, with most of the work occuring as part of short research and production residencies in different artistic institutions. Some points about the context of research-creation are worth mentioning in order to better understand the choices that were made.

First, the fact that we operated under tight budgetary constraints — at least in comparison with swarming robotics projects of comparable scope in science or engineering labs — had a tremendous impact on the components we chose to equip the bots with. In general, we favored cheap solutions that "do the job" and whose imprecision nevertheless added unpredictability to the piece, giving robots a wider range of behaviors. Simon Penny expresses a similar idea when talking about the "under-engineering" of his work *Petit Mal*, explaining how his approach of favoring cheaper solutions which are "70% reliable" over more expensive ones that might be "90% reliable" actually expands the "field of possibility" (Penny 2000, 401) by adding noise to the system, thus giving it more personality.[2]

These budgetary constraints forced us to carefully choose each component installed on the robots in order to give them enough capacity to be able to achieve the artistic goals while keeping costs and complexity low. We mainly focused our choices towards components that would give our agents; (1) a sensorimotor "body" that allowed them to have some minimal awareness of their surrounding

[2]An important difference between our approrach and that of Penny is that we compensated for a lot of the hardware's lack of robustness through software, whereas Penny, whereas Penny prefers to work with the material constraints rather than trying to overcome them with algorithms. His main critique is that "fixing in software" actually reduces the range of possibilities to be explored artistically. However, *Vessels* differs from *Petit Mal* in a number of ways, the most important one being the number of agents: whereas *Petit Mal* consists of only one robot which could be fine-tuned with a hands-on approach, in *Vessels* the large number of robots introduces a lot of variability, thus the degree of robustness needed in the components is more critical, as fine-tuning each robot would be too time-consuming. One of the favorable effects of developing these software algorithms is that our team contributed a lot of open-source code to the community, as we favored simple, low-cost solutions over off-the-shelf products.

environment and also to move and avoid obstacles; and (2) a variety of media-generation means to express a wide range of affects.

Second, it is worth mentioning that we went through several mock-ups and prototypes to arrive at the current design, and that there still remained space for improvement. At each step, we have validated certain items and made corrections in interaction with the different project participants. For example, software development at every step would reveal problems related to hardware, suggesting changes in the electronics components and circuit in the next iteration.

This section gives an overview of the technical aspects of *Vessels* robots while drawing links to the artistic intentions and to practice. I explain the propulsion and steering system, light and sound components, environmental sensing system and the infrared messaging system.

4.2.1 Locomotion and Steering

The propulsion system was the first element that we explored when we started working on the project. The vast majority of current research on robotics focuses on ground robots that move on wheels, or on flying devices (i.e., drones). There are hardly any instances of projects involving robots that move on water or other sorts of liquid.[3]

During our first residency at the CFAT in August 2010, we started experimenting with different kinds of propulsion and directional systems. We came up with an air propulsion engine using a pair of computer fans on servo-motors.[4] However, the fans were not powerful enough to fight against the wind and water current, while their large size made the robots less hydrodynamic.

For our second prototype, we opted for an alternative design involving a pair of stationary underwater pumps, which proved to work quite well. We also provided each bot with a magnetic electronic compass, allowing them to readjust their propulsion to stay on course and thus avoid spinning — an idea which proved to be more difficult to implement than it looked, as we shall see. We added an extra infrared distance sensor to the robots to give them a better sense of their

[3]During one of our public presentations at LABoral in 2013, an engineer working for one of the major petroleum companies in Spain left me a business card. He was interested in our design as a way to explore pipelines and detect cracks and other weak points.

[4]We came up with that idea originally because we feared underwater propellers could easily get stuck in aquatic weeds.

surroundings, facilitating obstacle avoidance among other things.

4.2.2 Expressive Gear

Some of the onboard components are specifically there to provide the robots with ways to manifest their personality, state, emotions, etc. to spectators. The most evident of these components is the color module, activated through three (3) RGB LEDs that allow the agents to display various behaviors or states by means of colored light.[5] A special routine running in parallel to the main code allows smooth color transitions and oscillations, such as a rapid fluctuation between red to blue.[6]

The robots are equipped with another light-emitting component which adds to the range of expressiveness. After the first prototype, Samuel suggested that we add a series of eight (8) bright white LEDs on the periphery of the printed circuit board. Unlike the RGB diodes, these LEDs can be individually controlled using a shift register. The physical placement of these light sources on the board opens up another layer of possibility: for example, we could now use these white LED lights to give a sense of direction as the robots are navigating on water.

Additionally, robots are equipped with a separate circuit board with its own, on-board 8-bit microcontroller and a small amplifying and filtering circuit, allowing the robots to generate 8-bit sounds at 16384 Hz. As I explain in the next section, this *sound/environment board* is also responsible for measuring a specific environmental condition. The board allows real-time audio synthesis of different sound types. For example, we implemented some very simple sounds such as white noise, as well as a more complex sounds such as the one generated through a genetic programming procedure. In most cases, these sounds can be tuned using a series of parameters, thus making it possible to generate an even wider range of audio effects.

[5] Although there are many LEDs, they are all programmed to glow in the same color. We increased the number of LEDs in each of our prototypes to achieve a satisfying level of brightness.

[6] In other words, rather than being limited to choosing one specific color, it is possible to create simple color animations that run in the background.

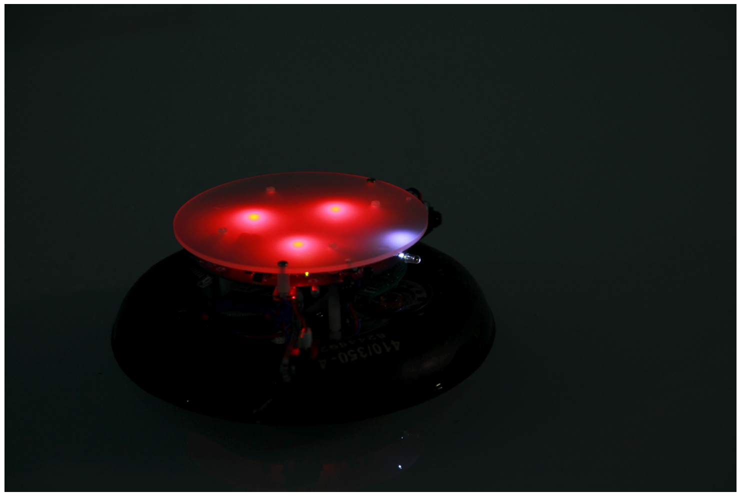


Figure 24: *Vessels* robot in action, showing the colored LEDs. Photo by Sofian Audry.

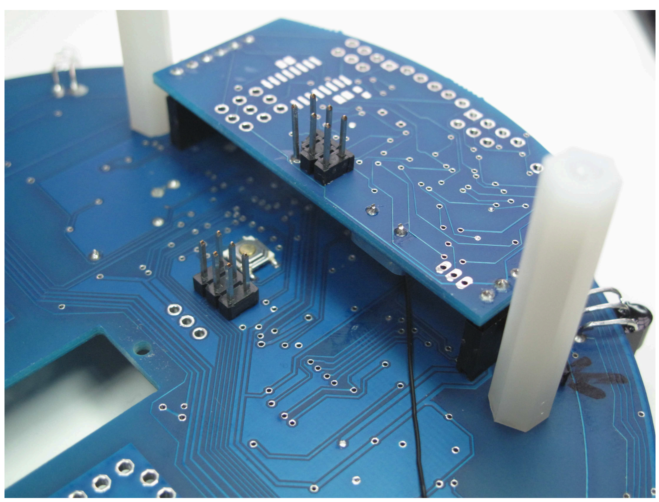


Figure 25: The sound/environment board connected to the main board (view from below). Photo by Samuel St-Aubin.

4.2.3 Environmental Sensing

One of the core conceptual elements of *Vessels* is the interdependence between the robots and their milieu. Each agent possesses a "piece of the puzzle", a personal view over its environment through a single sensor. These fragments of information are exchanged between the robots through their behavior, generating an emerging "rendering" of the data in the form of a collective adaptive performance.

Each robot is equipped with the exact same fundamental set of hardware and software (their environmental sensor is the only exception to that rule).[7] As each of these sensors can work quite differently and therefore need specific hardware and software components to function properly.

In our first prototype, we had to hard-wire sensors separately on each board, soldering specific circuit components and uploading the appropriate piece of code to work with it. This increased the risk of errors and made troubleshooting and maintenance more difficult. In our second prototype, we decided to create an external board which could be connected to the main one, a simple "plug-and-play" interface that hides the complexity and particularities of the environmental readings.[8]

4.2.4 Communication

Let us take a moment to reflect on the robotic bodies of the agents that I have been described thus far. A pair of water pumps are used to provide locomotion, colored and white LEDs for generating light effects, and an audio circuit for digital sound production. In terms of technology that facilitates the robots' perceptions, they are equipped with a compass for navigation, a pair of range finders for obstacle detection and a single, unspecified environmental sensor.

Now, notice that the range of perceptual data an agent has access to does not include movement, color, light nor sound. In other words, aside from the accidental case where the environmental

[7] In a given collective, there might actually be some sub-groups of robots who share the same kind of environmental sensor. There are two reasons for this.

First, it would be expensive timewise to research and test a different environmental sensor for each of our 50 robots.

Second, we actually believe it makes sense that some robots have the same kinds of perception, especially for cheaper components such as light and sound perception while having less robots equipped with more "special" sensors.

[8] As explained in the last section, we also thought it would be convenient to use the extra CPU power to run audio synthesis, which is why we integrated both components on the same external board.

sensor of one robot would be a microphone or a light sensor, these agents are unable to detect the behavior of their peers. Even if they did, it would be extremely hard for them to extract the complex information needed out of the flow of incoming data.[9]

Imagine that a robot starts acting in a specific way, like twirling, changing its color to green, or producing a chirping sound. How are its peers meant to react to these behaviors if they do not even have the means to perceive them?

We came up with a simple solution for addressing this issue. Instead of having agents notice what one of their peers is doing through a sensory system, we simulate the act of perception by having the perceived robots send the information about their behavior or state of being through infrared (IR) messaging.[10]

IR messaging systems are extremely popular in scientific swarm robotics applications for various reasons (Kornienko and Kornienko 2011, 9–10). One of the main advantages of using IR instead of radio frequencies (RF) is that it is possible to detect where the direction the message is coming from, allowing agents not only to exchange data but to locate one another[11].

As such, we equipped each robot with a series of infrared receivers installed around the perimeter of a simple technique described in (Hoyt, Mckennoch, and Bushnell 2005) which basically compares the intensity of the signal received from each sensor to estimate the angle and range of communication (Fig. 26).

There are, however, important limitations to using IR. Data transmissions are slow, short-distance and prone to error. This is accentuated by the presence of other sources of infrared light, with the consequence that we cannot show the work during the day due to interference from bright sunlight.

Another caveat is that the low transmission speed of infrared messaging increases the risk of message collision, which forces us to limit the frequency at which robots send signals. This is true

[9]Early on when, we started reflecting about the question of how the agents could socialize with one another, we considered the idea of using a simple, physical modality, like sound. For example, the robots could have reacted to audio signals coming from their peers or the audience. It could have been an interesting venue to pursue, but if we did so we would never have been able to achieve the level of interactional complexity we wanted to have in the piece.

[10]For example, instead of having robot A looking at robot B and seeing it spin clockwise (which would only be possible through a complex system of camera doubled with pattern recognition algorithms), robot B would send a message to A that says "I am spinning clockwise".

[11]It is also possible to guess the proximity of the sender using the infrared intensity of the signal.

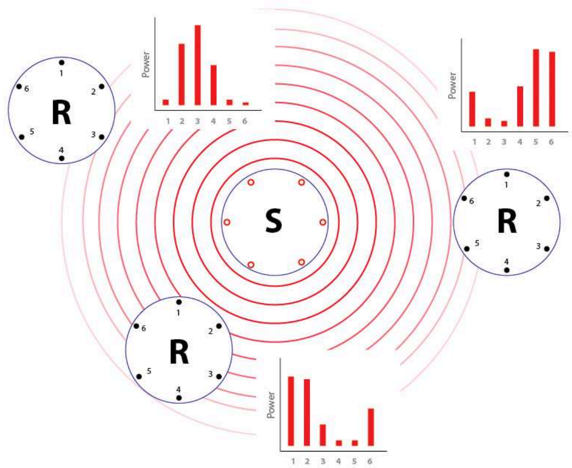


Figure 26: Diagram of the infrared guidance system. The sender robot (S) emits an infrared message in all directions. Receiver robots (R) perceive the signal through their six (6) infrared sensors with various strength. The relative signal intensities perceived by these sensors is used to estimate the orientation and distance of the sender robot with respect to the receiver. Image courtesy of Samuel St-Aubin.

for any application using IR, however the technology that we use is many times slower (2.4 kbps) than what is normally used in scientific research. As a comparison, the ISIS communication system for swarm robotics applications described in (McLurkin 2004) runs more than 100 times faster than our system at 250 kbps. Depending on the quantity of robots in the space, we thus need to limit the number of messages sent by a robot to about one every 20 to 30 seconds.

Distance also has an impact on the signal. In our tests, we could easily communicate up to 2 meters, but could hardly get to more than 6 meters while achieving the same results. However, we eventually came to see this apparent limitation as bearing a strategic advantage, as it constrained the agents in only communicating with their closest neighbors. We even lowered the IR intensity to limit the range further, bringing the outer limit of the signal to about 1.5 meters. Thus when a robot calls for peers to accomplish a social act, the radius of action of the signal naturally generates a local sub-group by being limited to the sender's immediate neighbors.

Quite interestingly, these technological limitations have forced us to address robots communications in a way that is actually much closer to the reality of biological systems. Living beings communicate using channels with relatively low bandwidth and range such as sound and gestures. The constraints hence coincide with our original goals for enabling IR messaging as complement to robotic action, which makes these actions perceptible to their peers. A faster, more expensive IR communication system would have allowed our robots to send hundreds of messages per second, but these messages would have been completely abstracted from the audience. In *Vessels*, this slowness means that each message accompanies a slow-paced, palpable behavior, such as emitting a particular sound or moving in a specific way, which brings the robotic communication to a "human level".

4.2.5 Control Unit

The behavior of agents is controlled from a central ATmega1280, an 8-bit microcontroller running at 16 MHz. We selected this technology for its small cost, low power consumption, high versatility and compatibility with the Arduino library[12], a very popular open-source software suite among artists and hobbyists.

Every piece of hardware is connected one way or another to this processing unit, and therefore have a software counterpart in the source code. A large proportion of the code is meant to provide a high-level interface to each of the subsystems to make them more robust and easier to work with.

The chips are fully programmable in a variant of C++, an extremely widespread and powerful object-oriented language. As such, it allows us to integrate nearly any kind of algorithmic techniques and structures, at least in theory. However, in practice, their low memory (8kB SRAM) and speed makes some applications difficult or even impractical, which excluded the use of some Machine Learning methods.

To summarize, the agents in *Vessels* are equipped with a range of low-fi hardware that "do the job", while being relatively imprecise when compared to more expensive commercial solutions. This lack of robustness is partly compensated for through software, but leaves room for more unpredictable behavior in the robots.

[12]The ATmega1280 corresponds to an Arduino Mega board. More information at: http://www.arduino.cc/en/Main/arduinoBoardMega

Whereas some contextual elements have been introduced through this short technical overview, in the next section we dig deeper into the research-creation process in an effort to address questions related to the practical aspects of using multi-agent systems and Machine Learning techniques in *Vessels*.

4.3 Practice

Through this technical overview, we have started to look at questions of practice, though only on the surface. Among other things, we have not really considered yet where the notion of adaptation fits in the project. I now want to start looking more closely at practice, describing in more details some aspects of the research-creation through three important phases of the project.

First, I examine a series of computer-based simulations of the work that were designed during Summer 2012. The simulations tried to give a sense of the global, emergent effect of robots moving over the water, directly addressing the artistic goals.

Second, I look at the way the control loop was implemented in the system. I explain how we started with Finite State Machines (FSM) in the first prototype and switched to a high-level goal-oriented model known as Behavior Trees (BT) in the latest prototypes. I describe the experience of working with these different approaches in terms of authoring and meeting our goals.

Finally, I review two ways we worked with adaptive systems: an approach based on genetic algorithms which was successful; and another based on reinforcement learning, which was not. I explain the reasons that have led us to experiment with these approaches, what the caveats were, and the final resolutions.

4.3.1 Simulations

In Summer 2012, the progress on *Vessels* was very slow. We were working on a second prototype, but had much trouble finding an appropriate physical space to experiment with such large groups of robots — especially one where we could install a big enough indoors basin filled with water. It was difficult to even begin to understand what things such as "emergent collective behavior" or "distributed choreography" really meant without first-hand experience with the installation.

In an attempt to make progress on the algorithms that could drive the piece, I implemented a software model of the installation which would allow me to simulate experiments of the work in a graphical interface using a 2D physics engine. The sensors, communication system and locomotion capabilities of the robots, as well as various parameters such as surface friction and wind, could be simulated in an integrated environment. The objective was not to create a perfectly faithful simulation of the robots but rather to experiment with simple algorithms to validate general ideas.

These simulations allowed us to come up with a set of fundamental concepts and heuristics for driving the work. First, the introduction of a high-level tension between sociability and solitariness became a way to generate contrastive variations in densities across space. Second, an embryonic notion of personality in the robots evolved through both social encounters and confrontation with the environment. Third, negotiation and generation of subgroup behaviors lead through temporary emergence of "leaders" in the community.

Solitary vs Sociable Robots

The first problem that I approached using the simulation engine was to create a distributed choreography of agents moving through space, seemingly governed by emergent, antagonistic movements of assembly and dissolution. The goal here was to find a simple self-organizing mechanism for governing how the robots occupy space, such that different distribution patterns could appear (isolated robots, sub-groups, etc.)

The basic concept that I experimented exploited the desire of the agent to either assemble with peers or avoid them. Following this idea, a robot would always be in one of two states: *solitary* (i.e., seeking loneliness) or *sociable* (i.e., looking for companionship). In the simulated experiment, robots regularly update their peers about their current state and location by broadcasting messages. Simple rules dictate the actions of a robot being in either state, based on direction of incoming messages and objects detected using the range finder, as portrayed in Algorithm 1.

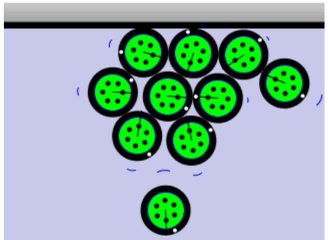
This simple algorithm inspires solitude-seeking robots to avoid objects, while sociable ones are drawn to one another. To prevent robots from staying indefinitely in the same state, I introduced an

Algorithm 1 Basic decision algorithm based on agent's sociability

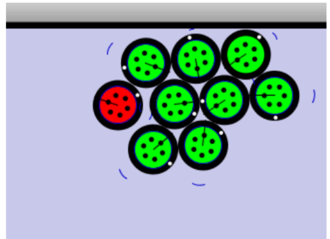
```
if agent is sociable then                                    ▷ seeking company
    if message received then
        steer towards incoming message
        start motor
    else if object detected then
        start motor                                          ▷ approaching
    else
        stop motor and steer clockwise                       ▷ searching
    end if
else                                                         ▷ seeking solitude
    if object detected then
        stop motor and steer clockwise                       ▷ avoiding
    else
        start motor                                          ▷ evading
    end if
end if
```

additional "boredom" parameter representing how much the robot has become tired of its current state, a value between 0 and 1 that slightly increases over time. The parameter is also influenced by the presence of other robots detected during the reception of infrared messages. When a robot's boredom reaches 1, it changes state and the parameter is reinitialized to 0. Robots trying to be alone will thus tend to become sociable through "peer pressure" if they move across a highly populated area, whereas groups of friendly robots sticking together for too long will irremediable get bored of each other after a while and split.

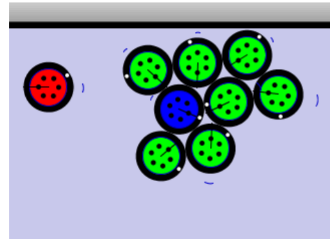
Once the experiment is launched, the robots start moving through space. Regions increase in density as sociable robots coalesce, while their individualistic counterparts roam freely, avoiding groups and taking refuge in deserted areas. Groups form, stick together for a while, then slowly split and dissipate. As they disband, the territory becomes more chaotic as it fills up with runaway robots running in all directions. Through their exploratory motion, these rogue robots reconfigure the space as they attract friendly robots, inadvertently bump into groups, and eventually switch states again and begin to seek new companions (fig. 27).



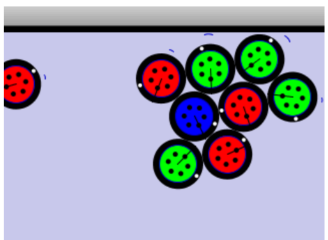
(a) A group of robots assemble.



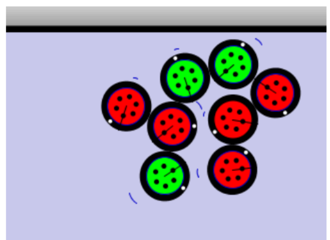
(b) One of them becomes bored...



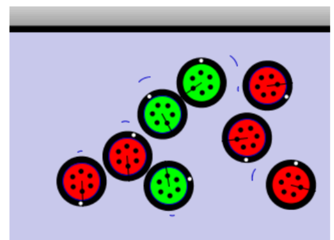
(c) ... and leaves.



(d) More robots become bored and leave.



(e) Runaways try to find a way out.



(f) The group dissipates. Loners will roam freely for a while until they change state and try to join other groups.

Figure 27: Images from the simulation engine: a group of robots splits and dissipates.

Tying Environment to Identity

These early software simulations revealed that complex movements of assemblies and dissolutions are possible using relatively simple rules, resulting in dynamic motions across space, generating various densities and often surprising events evocative of affective social tensions between agents. However, the behaviors were still limited because all robots behave and look the same: they seemed to be missing some kind of personality.

In an effort to address our artistic intention of generating an evolutive group behavior which specifically responds to the environmental conditions, I designed a new experiment using a fictitious environmental sensor that takes a measurement of the local environment such as room temperature or atmospheric pressure.

The scenario is amended as follows. Each robot is granted an *identity* parameter with a value between 0 and 1, a minimalist representation of the "personality" of the robot, initialized randomly.

Values near 0 and 1 represent more extreme, marginal identities, whereas "normal" is implied around 0.5 (50%). Robots with similar identity are deemed to be alike and tend to tolerate each other more, though they also get bored more easily if their resemblance is too strong. Conversely, robots with very different identities will find it difficult to stick together, though they might also be searching for that difference.

Instead of using color codes to show the current state of the robots, I choose to use the simulated RGB LED to represent the identity of the robot using a simple hue conversion. Agents near the middle point (0.5) are cyan (180° hue) while extreme values (near 0 or 1) both tend towards red (Fig. 28).

Figure 28: The hue scale, as used in the simulation engine to represent a robot *identity*.

Each robot is provided with an environmental sensor which also outputs a value between 0 and 1. The basic, individual behavior of robots related to their current state of being (solitary or sociable) remains unchanged from the last experience. However, the process governing the evolution of their boredom is now affected by both their identity as well as their environmental readings.

When a robot chooses to broadcast a message, the first thing it does is to update its identity based on the value of the environmental sensor using a simple adaptive rule:

$$identity \leftarrow identity - \eta(identity - environment) \tag{1}$$

One will recognize here, the general form of the negative feedback, stochastic learning rule introduced earlier, with a tunable learning rate η that can be adjusted to make the convergence slower or faster (it was set to 0.1 in the simulations). This rule allows the smooth, asymptotical convergence of the robot's identity to match that of its sensor. One way to understand this process metaphorically is that the value *environment* represents itself an identity that becomes the target of the robot, meaning that if the agent was alone and if the measured variable would never change over time, the agent would eventually stabilize its identity to match that of its perception.

Once its identity is adjusted, the robot sends a packet containing its new identity value. Upon receiving it, neighboring bots update both their boredom and identity parameters in ways that reinforce unions with peers of akin temperament. One of the important consequences of this procedure is that robots that assemble will tend to adapt their identity to one another using a learning rule similar to equation 1. As members of a party exchange messages, their identity parameter tends towards the group's average. Since the individual identities of the bots are in turn strongly affected by their environmental sensor, the robots will lean globally in the direction of the average output of the sensors.

For instance, if sensors are strongly biased toward zero, in the long run, the agents' color will be closer to orange and red. However, even if the mass effect tends to move in a certain direction, the spatial distribution of identities is subject to local variations. Thus, even if the global color converges to red, there will still be variations at the individual and group levels. For example, a party of robots with a higher identity parameter might gather in a safe spot, converging to blue or purple.

This basic mechanism, allowing inter-robot personality exchange through mutual adaptation to each other and their environment, would become the core principle behind the implementation of collective, emerging behaviors in Vessels. As we will see in section 4.3.2, the simple principle of mixing hues will later be augmented by the means of a genetic code in order to represent other behavioral components such as sound and movement.

Leader Bots

Thus far in the experiments, intentional movement was mostly reserved to solitary agents, while sociable robots stay more or less static once gathered in order to stick together. In a final experiment, I attempted to occasionally break from this principle by introducing a concept of leadership. From time to time, when a group is formed, a leader stands out and directs the other group members to swarm with it in its explorations of the space.

The last simulation adds a new layer to the "lonelies-friendlies" duality by enabling groups to

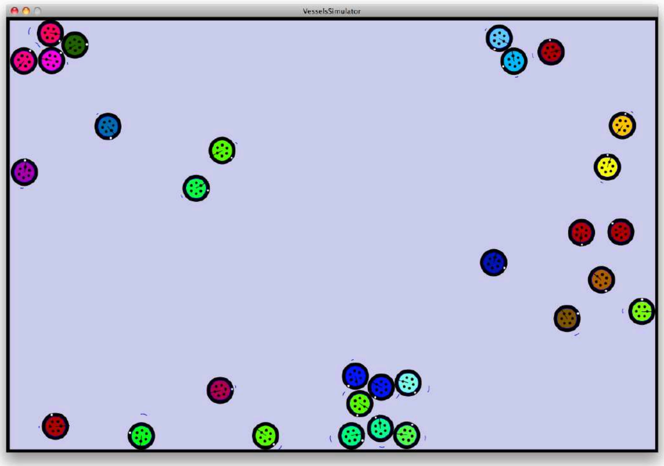


Figure 29: Screenshot of the simulation. Gathered robots tend towards a common identity. Solitary agents are represented with a darker color.

move in an exploratory fashion. As the notion of “leader” carries with itself a strong image of social systems, I made sure that no distinctive features really allow us to recognize the leader, so that the viewer might interpret these events in alternative ways, such as the singular desire of a group to start moving in unison.

The simulation experiments described here constituted a first step in the algorithmic development of large-scale collective behaviors. They allowed an efficient exploration of certain ideas using an iterative and incremental approach. One of the effects of these results was to give a sense of what the actual project could look like before actually spending all the time and energy to design the physical robots. It thus acted as a kind of rubber-stamp for me and my collaborator, validating our yet-abstract conception of what a large quantity of social robots could achieve aesthetically.

As we shall see, many of the concepts and high-level processes introduced in these experiments — such as the movements of unions and separations, the notion of identity in relation to the environmental sensors and the robots’ color, as well as the idea of leaders — have been preserved in the final source code running on the actual robots, although the final programs are definitely much more complex, partly because simulations could not foresee the intricacies of a real-life setting.

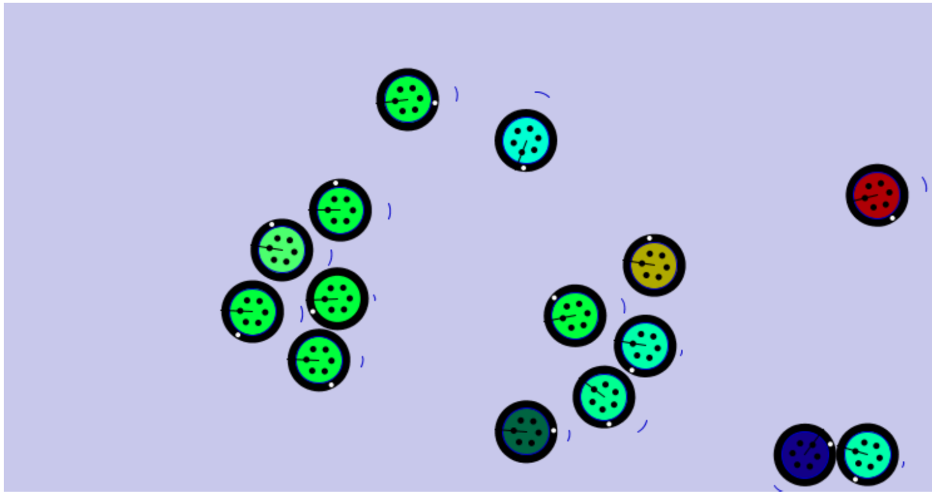


Figure 30: Screenshot of the simulation. The group on the left (in green) is led by a leader to move West.

4.3.2 Machine learning

We now enter in the core section of this chapter, where I will discuss our attempts at using Machine Learning methods in the process of implementing real-time behaviors in the robotic systems. It is important, however, to outline that the work does not rely uniquely on adaptive methods. The robotic behaviors actually consist of two components: (1) a formal, rule-based, goal-oriented system known as a Behavior Tree that runs the low-level decisions such as avoiding obstacles or fostering socialization; and (2) a custom modification of a Genetic Algorithm that is used to generate personalized manifestations of behaviors.

The Behavior Tree (BT) is a tree-like, agent-based structure that provides a high-level, stable interface that manages priorities and sequences of actions. Originally developed in the field of video game design as a surrogate to Hierarchical State Machines (HFSM) in the modeling of agent behavior (Isla 2005), it has been applied recently to robotic control (Marzinotto et al. 2014) where it can be considered as a flexible alternative to subsumption architecture.[13]

[13]Subsumption architecture is an approach to robotic control developed by Rodney Brooks in the mid–1980s, in response to some of the inadequacies of GOFAI. Whereas traditional AI approaches to robotics entailed creating an internal representation of the world, subsumption architectures directly couple observations to actions in a layered control system. Using a bottom-up approach, the engineer iteratively adds control layers, starting with the most low-level rules and refining them while moving into higher-level control routines. (Brooks 1986)

There were two things we wanted to explore with ML methods. First, we wanted to train robots to maneuver across the surface of a water environment, efficiently avoiding obstacles in an effort to generate aesthetically compelling motion. I tried implementing this using reinforcement learning, an approach that we finally abandoned.

Secondly, following our artistic objectives, we applied genetic algorithms to generate different forms of behavior that could evolve through time and be tied to the environmental readings. This approach was quite successful, although it came with some limitations which I will discuss further.

Governing Towards a Dead-End

One of the biggest technical challenges that we faced with *Vessels* was how to efficiently navigate the robotic agents through space. Their round shape and small weight combined with the low viscosity of water made them extremely unstable and naturally likely to spin, sending them quickly off course. Using compass data, I tried to design an algorithm that would allow them to complete the seemingly simple, basic task of moving straight forward in a given direction.

I started with the most straightforward program I could come up with, which basically attempted to readjust course by pushing the motors towards the target heading. In essence, the algorithm says: "If you are currently bearing left of target heading, move clockwise. Otherwise, move counter-clockwise."

This approach would work relatively well if the robot were starting with a heading close to the target heading. Otherwise, it would start making strong oscillations before adjusting or, in some cases, start spinning indefinitely.

The main issue here is that this approach does not take into account the angular velocity. If the robots need to move clockwise, the algorithm had better check whether the robot is already moving in that direction before pushing its motors to accelerate even more, making it harder to slow down later on.

I thus designed a simple mathematical model of the physics involved and used it to implement a new algorithm. It worked much better but it demanded that a number of parameters be set by hand through trials and errors, such as the impact of the motors on acceleration.

At this point, I started contemplating the idea of using reinforcement learning to automatically learn the system dynamics. This was an interesting candidate for RL because it is a clear optimization problem: the robot simply needs to be rewarded for keeping its current trajectory as close as possible to the target. The learning algorithm would do the rest, finding an optimal way of steering in any given circumstance. This could potentially save a lot of time and achieve better results as compared to fine-tuning a hard-coded program by hand.

While "moving straight" does not seem a particularly compelling aesthetic goal, I saw it as a first step in generating more complex and interesting behaviors in the robots. My hunch was that we could use the same approach to train robots to move across the surface to achieve different objectives, such as avoiding all obstacles, running into them purposely, maximize movement, etc., simply by changing the reward function.

At the end of our residency at LABoral (Gijón, Spain) in Summer 2013, I managed to train a system to move straight using batch RL (Lange, Gabel, and Riedmiller 2011). While those results seemed promising, when I started working again on it a year later during our residency at Eastern Bloc (Montréal, Canada), I found myself unable to reproduce them. All of my experiments yielded poor performances on the task, usually leading to a dreaded spinning behavior. I still do not fully understand what happened. My main hypothesis is that the environment we were in seemed to have unusual magnetic properties, probably due to large metal beams which supported the ceiling. This is demonstrated by the fact that as the bots moved on the surface, the magnetic field seemed to change the response of the compass, providing unreliable data which might have impaired learning. I had already spent way too much time working this out, and in the mean time had come up with an even better analytic solution to the steering problem which worked satisfactorily. As such, my research in RL came to a halt, and I began to concentrate on other issues.

Of course, it would probably have been possible to train the robots in this task given better conditions and more time. But we would then probably only have ended with similar results than what analytic solutions such as the one we implemented allowed, and it would have come at a much higher cost in terms of memory and CPU power. Furthermore, while the robots would have moved more efficiently, that would also have removed some of their clumsiness, which gives them more personality. In other words, indeterminacy played in our favor in the aesthetic design of the work. Again, this shows how the stakes in art and AI differ from one another and often even play antagonistically.

Nevertheless, as I mentioned earlier, learning the steering behavior was only a first step to open up a wide range of learned behaviors for the robots. But in the end, what would we gain from it, really?

Answer: probably not much.

The best we could hope for would have been to generate a surprisingly "intelligent" behavior for our robots, possibly evolving through time. But the truth is that (1) the physical limitations of the microcontroller we used would likely have prevented or at least impaired it; (2) the necessary time involvement was hard to estimate but possibly quite high; and (3) we could create something fairly close to this result, and possibly with more aesthetic success, using less resource-consuming methods.

This example once again points to an important limitation of ML (and of AI in general), which is that these approaches are directed towards problem solving and optimization, whereas in creative applications, the optimal solution is ill-defined. There is no such thing as the "best" joke, an "optimal" song or a "perfect" robotic behavior (Pasquier 2015). For instance, in *Vessels*, our objective was not *really* to get robots to move on water without hitting one another: what we are truly interested in was to generate a form of expressive behavior that matches our aesthetic intention.

The "optimal" shape of an artwork — if such a thing is indeed possible — is an extremely abstract concept even in the eyes of its creators. Were we to adopt an extremely naive conception of the creative process, where an artist would have an *a priori* vision of what she wants to accomplish and then materializes it in a work of art, that vision would necessarily be vague and difficult to express mathematically.

Of course, the reality is that artists usually do not know in advance exactly what they want to achieve. Artmaking is an intricate process, an ongoing interaction with material agencies that generally lead us to place we might have not anticipated. The "optimal" form of a work of art is often revealed by accident: you know it only when you see it.

The maladjustment of traditional AI to creative applications has been raised by a number of researchers in the field of digital creativity, a sub-field of computer science that tries to address the question of how we can make computers creative (Boden 2004). One of their critiques of AI is precisely the fact that it has traditionally focused on a certain category of problems (i.e., those that can be optimized) leaving aside a vast range of what the human brain can do:

> Artificial intelligence addresses the problem of emulating intelligence by having the computer achieve tasks that would require intelligence if achieved by humans. These tasks are usually formalized as well-formed problems. Rational problem solving is then evaluated by comparison to some optimal solution. If the optimal solution is theoretical and not attainable, optimization and approximation techniques can be used to get closer to the optimal, or at least improve the quality of the solution according to some metrics. Computational creativity is faced with the dilemma that, while creative behavior is intelligent behavior, such notions of optimality are not defined. (Eigenfeldt, Burnett, and Pasquier 2012, 144)

Reinforcement Learning is particularly challenging in artistic works because it is caught in a paradox. Generating aesthetically compelling content is usually a difficult, highly non-linear task which thus requires a lot of data and capacity (i.e., number of parameters/weights) for learning to occur.[14] But artistic contexts rarely offer these conditions. In the case of *Vessels*, for example, while data can be more or less easily generated by having robots traverse the water, the information was made too unstable to use because of the low-fi nature of the sensors used. Furthermore, the hardware running the algorithms was simply not powerful enough to handle it.

To summarize, in this first step, I faced an important problem related to the nature of AI research, which is that interesting artistic problems are intrinsically difficult for these methods to address, and at the same time, the kind of problems they are designed to solve are often not aesthetically interesting in themselves. In the next section, I present how we successfully subverted a genetic algorithm approach to generate a compelling evolution in behaviors by employing a set of strategies, introduced in section 3.2, which consists in utilizing the model and the evaluation function of a ML algorithm of an alternative way.

[14] Consider for example the curse of dimensionality, introduced in section 3.2.2 .

Evolutive Family Resemblance with Genetic Algorithms

As my experiments with RL proved unfruitful, I switched my attention back to the larger picture. By this point, we had implemented a basic, nonadaptive behavior that allowed the robots to occupy space dynamically and engage in basic social interactions using behavior trees. There were still important challenges to address. How can groups of agents evolve behaviors that converge to some form of similarity while staying diversified? How to make these behaviors correspond to the environmental conditions sensed by the robots? Moreover, how to accomplish this with a decentralized control where each bot makes its own decisions?

Simulations drafted a very rough sketch of how to attend to these questions by introducing a minimalist concept of "personality" or "identity" in the robots that would be adapted through the interaction with peers as well as readings of the environmental data. This factor, however, remained at the proof-of-concept level, only affecting the robots' color. But how to go further and allow for a large variation in personalities while preserving the form of co-adaptation experimented in the simulations?

Introduced in section 3.1.7, Genetic Algorithms (GA) offered an inspiring and useful framework for undertaking these challenges. The fundamental principles behind GAs were put forward by computer scientist John Holland in his famous study on adaptive processes (Holland 1992). His conception of adaptation is highly Darwinian, anchored in a certain idea about the evolution of the species through the survival of the fittest. Holland's contribution was to provide a mathematical framework for this process, turning it into an optimization method.

One way GAs have been utilized by artists is by tweaking the fitness function to match it to aesthetic preferences, a procedure known as Interactive Genetic Algorithm (IGA) (Dawkins 1986). Karl Sims' 1997 installation *Galápagos* is emblematic of this approach in new media arts. In this work, visitors can look at twelve (12) screens that display artificial 3D life forms. They can push a pedal located in front of each screen to select their favorite creatures, and a new generation of virtual beings is created based on that input.

Sims therefore lets visitors act as the fitness function: the GA is used directly to try learning "user preferences", whatever that means. Hence, despite its innovative character, I always found this piece to be quite didactic and artistically limited. Good art is not about providing people with a chance to confirm their preexisting tastes and preferences, but about bringing them into an experience that will surprise, move and, hopefully, transform the way they feel about the world.[15]

What I want to focus on here is another fundamental property of GAs, which is the ability to transmit genes from one generation to the next. This is typically accomplished by a crossover operator which takes the chromosomes from the parents and recombines them, producing offsprings that share genetic material with their progenitors. In other words, the recombination step (which is normally used as part of an optimization procedure) can be seen as a way to exchange genetic patterns among members of the population, resulting in the evolution of a "family resemblance" between them as they reproduce with one another. Thus, if we tweak or remove the selection step, we can end up with an algorithm that is no longer trying to optimize anything, but rather is furthering circulation of genetic information among agents.

This is what we did in *Vessels*. The idea is that each robot's personality would be represented not as a single real value — resulting in a color hue as its phenotype — but as a binary genetic code that would define every aspect of the robot's behavior and appearance. The same would also go for environmental measurements, each measurement being associated with its own virtual DNA (one way to see this is that each environmental measure has its own personality or identity).

When an agent encounters another agent, it mixes its own genetic code with its peer's, with the intention that its DNA moves a little bit towards that of its sibling; the same goes when it performs an environmental reading. We achieve this by performing one step of the genetic algorithm and selecting only the offspring whose DNA is the closest to the original code by defining our fitness function as the number of bits differing between the offspring and the original robot's DNA.

In other words, we hooked into the procedure and subverted it. We took took advantage of the particular qualities of the *model* (the binary DNA stream), removed the selection part of the *optimization procedure*, and used an *evaluation function* that would fit our aesthetic needs. We

[15]The installation *Performative Ecologies* (2008–2010) by architect Ruairi Glynn employs GAs in a similar fashion as Sims but in a much more experiential, less didactic way. In the work, which won the VIDA 11.0 award for art and artificial life, dancing robots compete for the audience's attention. Their behaviors are implemented using GAs. The fitness function here is the attention level of the public during each robotic performance (Glynn 2008).

thus tweaked the standard approach — whose aim is to converge to an optimal solution among the space of possible genetic codes — in order to use two features of the algoritmic *process* that we were interested in, namely (1) the recombination operation that promotes "family resemblance" between the bots, attached to the environment measurements; and (2) the iterative nature of the algorithm, which makes a form of narrative of adaptation unfold before the eyes of the audience.[16]

The manner in which to use ML in artistic contexts is not obvious, as most of the scientific research in the field has focused on applications whose needs are much different from those of artists. It does, however, provide interesting tools that allow the implementation of conceptual and formal ideas. In particular, adaptive computational strategies were used in *Vessels* to respond to the artistic intend of designing a collective and evolving emergent behavior that attempted to match, or tend towards, a set of environmental variables (which were potentially in constant flux).

But after considering all these practical questions, how do people react when confronted with these behavioral patterns? What kind of agency do they attribute to these robots, both at the individual and group levels? Does the use of adaptive and ML methods change anything in regards to an uninitiated human's perception? From an aesthetic perspective, how do these behaviors differ from other, nonadaptive emergent processes?

4.4 Ecosystems

Vessels is an instance of an approach that has been successfully applied over the years that involves bringing a large number of agents into an environment. This is reminiscent of the kind of artificial ecosystems projects carried out by ALife researchers such as Thomas S. Ray and his *Tierra* platform. Australian artist Jon McCormack is one of the leaders of this approach. One of his most important works, *Eden*, is an "evolutionary sonic ecosystem" that represents agents on a

[16] In addition to that general "personality-evolution" process, an important component of the sounds produced by the robots used another form of genetic learning. Stephen Kelly, who created this part of the program, used an evolutionary approach known as Genetic Programming (GP). A variant of Genetic Algorithms, the DNA of a GP system encodes sequences of instructions, which usually involve simple, low-level manipulations of memory registers.

Kelly used that system to manage the evolution of simple sound parameters controlling oscillators, such as amplitude and pitch. This results in a diversity of sounds that stay in the same recognizable range, giving an identity to the whole piece while allowing sufficient divergence to introduce novelties. In the most extreme cases, some robots adopt really bizarre, loud voices, or shrieking cries that range from adorable to annoying.

two-dimensional lattice, in a similar fashion as cellular automata. The agents react to one another and to their environment using a set of rules that are encoded as binary chromosomes, which are then evolved using a Learning Classifier System (LCS), a Machine Learning technique invented by John Holland with close ties to both reinforcement learning, supervised learning and genetic algorithms (McCormack 2009; Holland 1992; Urbanowicz and Moore 2009).

The work takes the form of an audiovisual installation in which the agents and their environment can be experienced as they move, mate, eat and communicate with one another using sound signals. Contrary to Karl Sims' *Galápagos*, in which the audience is asked to directly influence the evolutionary process by acting as the fitness indicator, here the agents react only indirectly to the visitors whose presence is necessary to add "food" to the environment, while the movements of the audience increase the mutation rate (McCormack 2009).

As for the public reception of the work, McCormack notices that while most people "are not aware of the learning system, camera sensing, even the fact that what they are experiencing is a complex artificial life system", the system does seem to "interest and engage the audience":

> Since the system is reactive to people (rather than interactive), there is no correct or incorrect way to behave except to appreciate the experience. Anecdotal accounts from people who have experienced the work describe it as "having a sense that it is somehow alive," or "like being in a strange forest at night." In a number of exhibitions, people returned to the work over a period of several days, to see how the qualitative behaviour of the virtual environment had changed. In one recent exhibition, a local businessman visited the work during his lunch-hour every day for 3 weeks, describing the experience as "fascinating... one that made me more sensitive to my own environment." While these are, of course, subjective evaluations, it does appear that *Eden* is able to adapt and evolve to create an ongoing interest for its audience. (McCormack 2009, 411)

Performative Ecologies (2008—2010) by architect Ruairi Glynn, is described by its author as "an ongoing investigation into the design of conversational (interactive) environments" (Glynn 2008). Inspired by the work of Gordon Pask, especially his 1968 installation *Colloquy of Mobiles*, Glynn's installation creates a conversational space in which dancing robots evolve in constant interaction with one another and with the public.

> The performances are generated from a gene pool of evolving dances functioning in

> a Genetic Algorithm (G.A.) which uses facial recognition to assess attention levels & orientation of the audience before & after each performance as a way of assessing & assigning a fitness value to each new choreography. Over time successful maneuvers are kept & recombined to produce new performances while less effective ones are discarded. Mutation in the G.A. fluctuates based on how successful the sculptures become. If they get a lot of attention, mutation levels rise as if they are getting arrogant & as a result be come more experimental. (4—5)

When no one is around, the robots start communicating with one another, sharing information about their most successful moves. New performances are evolved through genetic crossovers:

> They take the suggestions of their surrounding partners & compare their gene pool of performances to their partners suggestions. If they are comparatively similar then they are accepted & replace a chromosome from their own pool. If they are too different they are rejected as if they dislike the partners dance moves. (5)

Performative Ecologies somehow lies between situated robotics, ecosystems installations and interactive genetic algorithm systems such as Karl Sims' *Galápagos*. Building upon the legacy of artist David Rokeby through his "experiments in complexity" (3) and Gordon Pask's Conversation Theory (which suggests a way for humans and machines to interact within their shared environments), it uses Machine Learning as a way to engage the public in sophisticated interactions with machinic systems.

Another example of an artistic installation driven by a virtual ecosystems is Richard Brown's 2000 piece *Biotica*. In this one-person immersive environment, it is once again the presence and movement of the participants that impact the growth and evolution of virtual 3D creatures. The piece integrates different methods borrowed from AL and AI, such as systems theory, genetic algorithms, L-Systems and Kauffman networks. Interestingly, agents in *Biotica* are modeled using simple neural networks that are evolved with a genetic procedure (Brown et al. 2001).

In an honest self-critique of his work, Brown highlights important difficulties in the research-creation process. In particular, he explains how the system interactions often resulted in chaotic behaviors that were too random, therefore impossible to stabilize. There were some "promising glimmerings" of emergent properties but in the end, in far too many instances, "the system had to be painstakingly coaxed, and behaviour had to be explicitely programmed to a degree that would not warrant the description emergent" (78).

In the case of *Vessels* neither the agents nor the enviromnent are simulated, as the piece stages a large group of situation robots that interact within their environment. In a similar fashion as *Eden*, the work uses genetic computation to evolve a group behavior that reacts to the characteristics of its milieu. A common feature of these works is that they presume a certain degree of loss of control from the artist, and their complexity cannot be easily grasped by the audience. How does the presentation context influence visitors' experience of such works? Do they require technical explanations to appreciate them? What are their first impressions, and how do these evolve as they spend time with the work?

4.5 Experiencing Vessels

At the moment of writing this thesis, *Vessels* has been presented on multiple occasions. In this section, I describe the experiences of the audiences in two of these showcases: one where the work was shown indoors as part of a gallery exhibition; and another where it was installed in the public space during a new media art festival.

In February 2015, a group of ten (10) robots was shown at the Eastern Bloc gallery in Montréal (Canada) as part of a group exhibition titled "Robotis Personae". The piece was thus presented *intra muros*, running during the space's normal daily opening hours. The reduced number of agents diminished opportunities of emergent phenomenon, and the gallery setting induced a certain expectation of reactivity that was not present. The context of an indoors exhibition did not favor the kind of detached, contemplative mindset the piece was designed for — which can be more easily found in people approaching an outdoors water feature in the city to eat a sandwich.

The most interesting perspective shared as a result of exhibiting this artwork came from a fellow media artist, Audrey Samson, who was doing a residency at Eastern Bloc while the exhibition was taking place. Because she experienced the piece on a daily basis, her experience was fundamentally different than that of most people who came to see the work in the gallery. She contacted me to tell me that, while she was unsure at the beginning, as she went pass the piece to get to the coffee machine, she became increasingly familiar with it and started noticing patterns. In particular, she explained that as a given day would go by, it seemed that the behaviors became less and less random and seemed to converge and stabilize into more recognizable, interesting patterns.

Several months later, we showed the piece at University of Strasbourg (France) as part of *L'Ososphère*, a public art festival running on the campus in November 2015. There, we exhibited twenty (20) bots in a public fountain set in front of the Law School, near a large and busy boulevard. The piece was shown for four (4) consecutive evenings and attracted passers-by and festival attendants alike.

In this case, the robots gave rise to a number of different social interactions and impressions in the recipients. One of the unexpected phenomena was that people, especially children, were immediately drawn to physically interact with them, either pushing them away, or even "testing" them with the hand or foot. We even had two cases were robots were flipped over, putting them into immediate risk of damage. I have mixed feelings about this: in one way, it got me exasperated, but on the other hand, it shows the different beliefs people have about AI, which range from curiosity and love, to fear and anger. In particular, the apparent fragility of these robots asks for a sense of responsibility in the face of their precariousness.

Children loved to help the robots escape from the sides of the fountain (a tendency they have due in part to their incapacity to move backwards) by pushing them away, usually gently enough that they would not tilt over. One night, two kids played for over half an hour with a single outlier robot who had run to the other side of the pool, apparently not caring much about the rest of the robotic community. It became a kind of game, and French philosopher Michel Serres' concept of quasi-objects immediately came to my mind. According to Serres, such quasi-objects are what "trace" relationships between members of a collective, like the ball in a soccer game. Through these interactions, they actually define the collective (Serres 1982).

People of different age and apparent socioeconomic and racial backgrounds made remarks that were more than often related to attributions of emotions and feelings in the agents. When asked about the installation by a group of baby boomers that had been observing and photographing the piece for half an hour, I responded by reversing the question, asking them what they felt the piece was about. They then started to describe all kinds of relationships that they imagined between the bots. For instance: there was a couple forming over there, these three were dancing together, this one was a loner, etc. We discussed whether one of the robots, who had got away from the rest, was alone by choice or because it had been left alone by his comrades.

On the last night, a 10-year-old girl arrived with her mother and seemed really intrigued with the piece. When told that the robots could communicate, the child asked whether the one receiving a message could tell that it was another robot who was contacting it. She also asked if a robot could tell another one that it was beautiful.

Many people (especially those who had just arrived at the site) asked whether the robots' behaviors were "random". I was surprised by this remark, because in my extensive experience working with the behavior of technological agents, I experience randomness — for example, the pattern generated by a random walk, where a completely random decision is taken at every step — as having a distinctive "feel" to it, which is quite distinct from that in *Vessels*. When asked what they meant exactly by that statement, many seemed to define "randomness" as oppositional to "programmed". What I understood by further discussing with them was that in their mind, randomness meant total chaos, meaning no perceivable pattern spatially and temporally; while a programmed behavior was to be naturally ordered and comprehensible, a set of defined steps, like a kitchen recipe.

Once explained that they were imbued with some degree of indeterminacy but yet had "personalities" and "desires", they often started looking at them differently, attributing more human characteristics. A visual artist told me that her perspective changed once she read the description of the piece. It seems that many people interpret such robotic systems using a mechanistic view. At best, the most informed people might think of a logical, rule-based, deterministic process. But until they are informed of the details of ML involved in the work, most people do not seem to be aware of the possibility that the robots might be adaptive, or emergent, that they might have their own agenda, that they might be shifting from chaos to order in performances that appear relatively seamless to the human eye.

Periods of social harmony in the robots contrasted with the clashing behavior of outlier agents. While the robots seemed to reach points of stability, most noticeably in the tint of their colored light, over time they evolved and transformed, not only in their individual expression, but also in response to the wind and water current which directly influenced their spread across the surface, which correspondingly influenced the distribution of spectators. In this way, the robots perhaps became a reflection of their human counterparts. A passerby even suggested that they might actually be the ones observing "our" behavior as the work of art.

4.6 Conclusion

In this chapter, we examined the agent-based installation *Vessels* in an attempt to address issues related to the use of Machine Learning in agent-based artistic installations. We started by discussing the authorial intentions behind the piece, namely (1) the dynamic occupation of space; (2) the generation of social interactions; and (3) the creation of a relationship between the robots and their environment. We then focused on the different hardware and software components of the piece, trying to contextualize these choices as part of the broader artistic research and development process. I explained how budgetary and human resources, time constraints, and goals of the enterprise differentiated it from most commercial or scientific swarm robotics projects, resulting in technological choices that favored expressivity over physical efficiency.

We then moved on to the crux of the chapter, looking directly at the storyline of some of the important practice-based research that was carried over as part of the work. We examined how high-level solutions were first considered through simple algorithms in a simulation environment. We inspected the diversified structure of the program which intermixes a nonadaptive, goal-oriented system with adaptive procedures.

I examined two strategies of using ML in the work. The first one, which tried to employ a RL approach to address the problem of moving efficiently across space, failed for both technical and artistic reasons. As I explained, this points towards the fact that, as other researchers have noticed, the trouble with using AI methods in art is that their goals mismatch. Indeed, AI is traditionally aimed at rational problem solving and optimization, whereas artistic applications have more murky objectives.

In response to this, I came up with a second stratagem using GAs, which involved hijacking the learning loop, using the iterative, hill-climbing nature of the algorithm to induce small adaptive changes that move towards a “family resemblance” over time, the objective of which was to allow learning to be “felt” by the audience over time.

From a broader perspective, my analysis points towards a way of using ML artistically in agent-based artworks by revealing the learning behavior itself to the audience in real-time, an antipodal approach to traditional AI, which is (supposedly) interested in an algorithm’s outcomes rather than in the process itself. I showed one very specific way one could hook into the optimization loop by tweaking the returns function (i.e., the fitness in the case of GAs and the reward in the case of RL).

Yet, the final use of ML in *Vessels* has many caveats. First, while we did approach the project from the perspective of ML, our reliance on Genetic Algorithms puts the piece in the space of evolutionary devices such as Artificial Life ecosystems, which we will discuss in the next chapter. Thus, it is possible that we could have come up with a similar system if we had an approach that was not explicitly based on ML but rather on self-organization and complexity.

Secondly, as the work already had too many degrees of freedom due to the water environment and the lo-fi sensors and motors, we ended up relying on a form of adaptation whose outcome was disconnected from the embodied reality of the robots. This is indeed what was attempted through the use of RL (but this approach eventually failed): to allow the robots to generate their own movements under constraints. Instead, the GAs are used here more as a content generation device that allows for similarities to emerge than as a way for agents to come up with their own, creative behavior.

Finally, the aesthetic effect of the robotic agents’ behavioral patterns, and how adaptiveness contributes to it, are still unclear. If we are to understand how people react to adaptive behaviors, we have to discern how they react to artificial agents in general, and what roles are played by notions such as autonomy, emergence, self-organization and adaptation in the generation of such artificial behaviors.

In the next chapter, I introduce a parallel strand of research to AI and ML which deals with questions of aliveness and embodiment using a bottom-up approach. I dig more deeply into the rich conceptual soil surrounding lifelike agents in art and science, showing the relationship of notions such as autonomy and emergence with adaptation and learning. Finally, I use the morphological differences between processes generated by adaptive and nonadaptive systems to construct a aesthetic framework for agent behaviors.

Chapter 5

Aesthetics of Behavior in Agent-based Art

> As observers we expect the environment to change and try to describe those features that remain unchanged with the passage of time. An unchanging form of events due to the activity within an assembly is called a behavior. The behaviour of a steam engine is a recurrent cycle of steam injection and piston movements that remains invariant. The behaviour of a cat is made up of performances like eating and sleeping and, once again, it is an invariant form selected from the multitude of things a cat might possibly do. The behaviour of a statue is a special case, for the statue is immobile, or to use an equivalent formalism, it changes at each instant of time into itself.
>
> – Gordon Pask, *An Approach to Cybernetics*

The first chapter of this dissertation explored the research-creation project *Absences*, a series of site-specific environmental works using electronic agents installed in outdoor spaces. The evolution of the project through the five (5) interventions that were realized demonstrated the origin of the current research project while also opening up several problems and questions.

In the second chapter, I uncovered part of the historical background surrounding the algorithms employed in *Absences*, and I introduced a framework for understanding the use of Machine Learning in art practice through an analysis of their innate properties as scientific objects.

The third chapter focused on the swarm robotics installation *Vessels*, showing an in-depth research-creation project that began with the premise that it would employ adaptive methods. It offered an opportunity to apply the framework built in the previous chapter, showing its forces and limitations. It ended by looking into the aesthetic effects such adaptive behaviors generated in a human audience, expanding this into possibilities for how audiences might react in future applications..

The current chapter follows from its predecessors in its attempt to tackle the question of affects and experience through a study of real-time behaviors in agent-based systems. In other words, it completes the theoretical framework of chapter 3, which focused more on the question of practice (i.e., the *artistic* dimension), by looking more closely at the question of experience (i.e., the *aesthetic* dimension). It does so by exploring the concept of behavior in art and science, using adaptation as a way to refine existing frames of reference.

In other words, this chapter presents an aesthetics of adaptive agent-based installations through a morphological analysis of their behavior. Following Kwastek (2013), the word "aesthetics" is used here as a fluctuating notion that ranges from "perception mediated by the senses" (aisthesis) to "theory of art" (aesthetics) (66). My main objective is to provide a description of the experiential mechanisms that are made possible by adaptive systems in media installations. I am especially interested in connecting the dots between the scientific perspectives of such systems and the aesthetic effects they afford. While artistic media installations cannot be separated from their visual and aural qualities, here I am interested in another dimension, which is about the *behaviors* that guide how these sensual elements manifest themselves in time, through the actions of an agent responding to its environment.

This chapter introduces a number of missing concepts in the history of adaptive systems that are necessary to understand my own aeshetic of behavior: Second-Order Cybernetics and Artificial Life (Langton 1989b; Helmreich 2000; Varela and Bourgine 1992; Maturana and Varela 1980;

Whitelaw 2004); the question of embodiment in enactivist theory and Nouvelle AI (Varela, Thompson, and Rosch 1991; Brooks 1987; Penny 1997); the notions of emergence and self-organization in agent-based systems (Cariani 1989; Soler-Adillon 2015); and the importance of authorship and believability in designing such systems (Bates 1994; Downie 2005; Breazeal 2002; Mateas 2001). I then contextualize adaptive behaviors within the scopes of analysis of social robotics through the work of Sherry Turkle (Turkle 2006) and performativity theory using Andrew Pickering's posthumanist ontology (Austin 1962; Pickering 2010). Following this, inspired by the work of Iannis Xenakis and Agostino Di Scipio on generative music, as well as on the writings of Cariani on adaptive and emergent systems, I suggest a simple aesthetic framework for understanding behaviors based on their morphological unfolding through time.

5.1 Second-Order Cybernetics

The historical review of chapter 3 followed a specific stream of Machine Learning, looking at its origins in Cybernetics and AI. However, in the artistic world, another strand of research that has ties with Cybernetics has been much more influential than AI: that of Artificial Life (AL, or ALife). This strand of research is marked by an interest in generative and evolutive processes rather than adaptive ones, and favors a "bottom-up" approach (as opposed to "top-down").

In the 1960s, some cyberneticians had started to distance themselves from the original movement of the 1950s because of an important philosophical limitation of the homeostatic model: the problem of the observer. The first wave of cyberneticians kept the observer outside the system, a shortcut that allowed for elegant mathematical modeling, among other things. But as Katherine Hayles puts it, the problem is that "feedback can also loop through the observers, drawing them in to become part of the system being observed." (Hayles 1999, 9)

Austrian-American scientist Heinz von Foerster called at the end of the 1960s for a new wave of Cybernetics by proposing to reinstate the observer in the homeostatic model by including it as part of the system. In other words, the cybernetician had to take his or herself into account for Cybernetics theory to be complete (Clarke and Hansen 2009). As the author would later put it, "Cybernetics then becomes cybernetics of cybernetics, or second-order cybernetics." (von Foerster 2003, 287). Thus, as Hayles puts it, in this second movement of Cybernetics, *reflexivity* took the place of homeostasis as a central concept.[1]

It is during that period that Chilean biologist Humberto Maturana and his student Francisco J. Varela put together the concept of *autopoiesis*, a reflexive and self-organizing model of the living, which quickly became extremely influential in the field (Maturana and Varela 1980). The link with homeostasis is clear: autopoietic systems *are* homeostatic, however, all homeostatic systems are not autopoietic. In autopoietic systems, the variable that is kept constant through homeostasis is none other than the organisms' own organization (79).

Maturana and Varela's main interest is to establish a common ground for living systems. They take, as the cornerstone of their theory, the unitary character of every living system. Although their approach is mechanistic, they emphasize that their interest is not "in properties of components, but in processes and relations between processes realized through components" (75). Autopoietic machines are closed, autonomous systems that maintain their individuality and their unity by a process of self-production that constantly builds and rebuilds the components that make up their organization.[2]

Thus, contrary to what Rosenblueth, Wiener and Bigelow suggest, living systems are not defined by their purpose, nor by their inputs and outputs: these features belong to the domain of the observer, not to the phenomenological domain of the autopoietic system.

> Purpose or aims [...] are not features of the organization of any machine (allo- or autopoietic); these notions belong to the domain of our discourse about our actions, that is, they belong to the domain of descriptions, and when applied to a machine, or any system independent from us, they reflect our considering the machine or system in some encompassing context. (85)

Another consequence of their definition is that although the authors recognize the link between autopoiesis and self-reproduction in the history of evolution (106), they do not consider

[1] This first-order vs second-order cybernetics dichotomy, taken for granted by Hayles and other scholars, has been criticized under both historical and conceptual grounds. For example, read Cariani (2016).

[2] German biologist Jakob von Uexküll's notion of umwelt can be considered a precursor to the autopoietic theory. Von Uexküll notes that animals are engaged in a functional loop where perceptions control actions that contribute to change the world they live in, which is completely different from the world of the observer (von Uexküll 1957).

self-reproduction to be a necessary condition of life. For a unity to reproduce, it needs to exist first and foremost. Thus, self-reproduction is "operationally secondary to the establishment of the unity" and, for the authors, should not be part of the definition of living systems (100).

5.2 Artificial Life

As second-order Cybernetics re-introduced the observer into the system, by removing self-reproduction as a central feature of life, it failed to take the evolution of living systems into account, especially in the context of complex, non-linear dynamics. Starting in the 1980s, a new approach known as Artificial Life — or ALife for short — developed in response to these caveats (Hayles 1999, 222).

In the 1970s, chaos theory and complex system theory had revealed how highly non-linear systems often display *emergent* properties, that is, unpredictable behavior as the result of simple interactions between a large number of entities. Emergence challenges directly the distinction between human and machine because we can now, starting from simple rules, simulate complex and unpredictable behavior on a computer.[3] Thus, in this third wave of Cybernetics, according to Hayles, reflexivity gives way to *virtuality.* To understand Hayles's definition of virtuality, one need only think about an immersive game where a human body is put in a feedback loop with a 3D simulation on the computer. The game thus happens at the intersection between the real and the virtual:

> Virtual reality technologies [. . .] make visually immediate the perception that a world of information exists parallel to the "real" world, the former intersecting the latter at many points and in many ways. (14)

It should have become quite apparent by now that this kind of computationalist conception that supposes a clear separation between the informatic world and the material world is a powerful thread in the history of computing. In particular, it is clearly manifested in Christopher G. Langton's opening opus to the proceedings of the first Interdisciplinary Workshop On The Synthesis And Simulation Of Living Systems at the Santa Fe Institute (Langton 1989b). In this foundational

[3] Emergence is a key concept of complexity sciences, such as chaos theory (Gleick 2008), which were very influentual on the upbringing of Artificial Life (for example, see (Langton 1990)).

paper (incidentally titled "Artificial Life") Langton frames the field of Artificial Life as a synthetic approach to biology.

Langton explains that biological sciences are anchored in an analytic methodology that tries to understand life by looking at the only forms of living systems that we know of: carbon-based life on Earth. This limitation to terrestrial examples of life makes it hard to derive general principles for biology. Artificial Life proposes to overcome these constraints by creating "life-like behaviors" within the computer. This new "biology of possible life" thus supplements traditional biological sciences with a synthetic approach: "By extending the empirical foundation upon which biology is based beyond the carbon-chain life that has evolved on Earth, Artificial Life can contribute to theoretical biology by locating *life-as-we-know-it* within the larger picture of *life-as-it-could-be*." (Langton 1989a, 1)

Although this broadened picture should theoretically encompass hardware-based as much as software-based systems — not to mention synthetic forms of carbon-based life, often called "wetware", inclusive of cyborgs and other hybrids — Langton gives little importance to the material support of living systems, whether it be sillicon chips, biochemical structures or mechanical devices. ALife, he explains, is rooted in the history of computing and more especially in the switch from a mechanical to a *logical* — in other words, computational — conception of life (13).

Langton's support of the computational approach can be seen in his appraisal of John von Neumann's theory of self-reproducing automata, the "first computational approach to the generation of lifelike behavior" (13). Von Neumann, a genius of his time, had been a key player in early Cybernetics and in the design of modern computers. In the early 1940s, with his student Stanislaw Ulam, he designed a theory of cellular automata as a computational model of living systems. A cellular automaton is a set of discrete cells arranged into a grid, each of which can be, at any given time, in only one of many states. Cells change state according to a set of rules that depends on the states of its immediate neighbors (von Neumann 1951).

Von Neumann was interested in the capacity for such an automaton to self-reproduce. He set out to design a cellular automaton capable not only of reproducing itself, but also any given automaton. Von Neumann called that automaton a "universal constructor". Unfortunately, he died before he was able to publish his research. It is Arthur W. Burks, a computer scientist who had worked on cellular automata in the 1940s and who contributed to the design of the first general-purpose computer, who in 1966 edited a posthumous publication of John von Neumann's research on self-reproducing automata (von Neumann 1966).

In a 1984 paper, Langton designed a simplified version of von Neumann's automata. The universality condition was dropped: the automaton should only reproduce itself, although it should do so in a non-trivial manner. The resulting structure consists of a dynamic "loop" that stores its own description. Reproduction occurs by the loop extending into an "arm" (ressembling an umbilical cord), which in turn carries the code for creating another loop (Langton 1984). These "Langton's loops" thus exhibit a life-like behavior that is reminiscent of the steps of a biological life cycle (143).

In his 1989 article "Artificial Life", Langton's bias towards the computationalist view of life displays itself not only in his depiction of ALife's origins in von Neumann's theories, but also in the examples he chooses to describe the field. Most of his attention is focused on purely computational models of life, like Lindenmayer's generative grammars that grow artificial trees (Langton 1989a, 25), cellular automata (28—30), Reynolds' flocking "boids" (30—31) and genetic algorithms (35—38). To his defense, Langton does venture into hardware-based artificial life in the post-war period, such as Grey Walter's electronic tortoises (18) but he clearly insists on the primacy of behavior over matter: "Life is a property of *form*, not *matter*, a result of the organization of matter rather than something that inheres in the matter itself." (41).

The *Tierra* system, designed by ecologist Thomas S. Ray in the early 1990s at the Santa Fe Institute, is probably the most emblematic example of this computational view of life (Ray 1991). In this experiment, Ray raises the question: what would life look like in the universe of a computer? To address this question, he establishes a metaphor in which processes reproduce, mutate and evolve within computer memory. He thus establishes a direct parallel between organic and digital life.

> Organic life is viewed as utilizing energy, mostly derived from the sun, to organize matter. By analogy, digital life can be viewed as using CPU (central processing unit) time, to organize memory. Organic life evolves through natural selection as individuals compete for resources (light, food, space, etc.) such that genotypes which leave the

> most descendants increase in frequency. Digital life evolves through the same process, as replicating algorithms compete for CPU time and memory space, and organisms evolve strategies to exploit one another. CPU time is thought of as the analog of the energy resource, and memory as the analog of the spatial resource. (Ray 1991, 373—374)

By running this program, Ray observes the occurrence of phenomena that he himself had not foreseen, such as the emergence of different forms of parasitism (383—387). Here I want to draw attention to the fact that, for Ray, the creatures that dwell in *Tierra* are real life forms. *Tierra* is not a simulation of life, but rather a ground for the emergence of digital life form, as he expresses very clearly in his introduction: "The intent of this work is to synthesize rather than simulate life." (111).

Thus, many ALife scientists share a computationalist vision of life, in which information subsumes matter. But as Hayles rightfully claims, others are opposed to this view. In his book *Silicon Second Nature*, Stefan Helmreich gives an anthropological account of Artificial Life research at the Sante Fe Institute (SFI) where he highlights the largely computationalist nature of ALife in the US. In the last chapter of the book, however, he contrasts it with a "European school of Artificial Life" that has its own historical track. Whereas the US/SFI school of ALife is usually associated with Langton, the European school is associated with Varela and reclaims its historical links with Walter and Ashby's cybernetic creatures, distancing itself from the computationalist tradition (Helmreich 2000; Varela and Bourgine 1992).

Like Cybernetics in the 1960s, the field of Artificial Life would open up a whole new territory for artists after this period of growth in the late 1980s. In his Ph. D. dissertation, new media theorist Mitchell Whitelaw attemps to define this area of artistic practice. He remarks that Artificial Life (ALife) is an area of experimental science which is less preoccupied by observation and representation than by by intervention and action. He also opposes it to the field of Artificial Intelligence (AI) which promotes a "top-down" rather than the "bottom-up" approach of ALife. Tracing through the interests of art in regards to synthetic life over time, in artists and thinkers such as Goethe, Malevich, Klee and Schöffer, he hypothesizes that "a-life art" might just be the latest addition to "a modern creative tradition that seeks to imitate not only the appearance of nature but its functional structures" by using or appealing to technology. ALife might then just be the true destiny of art and the realization of Jack Burnham's vision of a "living, cyborg art form". (Whitelaw 2004, 19)[4]

5.3 New AI

Influenced by approaches in both Machine Learning and Artificial Life, as well as by the work of Maturana and Varela (Maturana and Varela 1980), MIT robotics scientist Rodney Brooks challenged classical AI by proposing a "New AI" or *Nouvelle AI* at the end of the 1980s. According to this view, living systems should not be seen as the mere substrates on which a disembodied series of symbolic manipulation happens. On the contrary, Brooks proposed that the behavior displayed by living beings results from an embodied, situated interaction with their environment which does not have a need for intermediate representations of the world (Brooks 1987).

Brooks' robotic systems bear many resemblances to the works of early cyberneticians such as Grey Walter, who created in the 1950s a couple of electro-mechanical turtles that could accomplish complex, life-like behaviors through the use of a simple set of procedures that took into account their specific bodily attributes (Walter 1950). Brooks' subsumption architecture, which allowed him to create his first walking robot, *Genghis*, displays learning capabilities and has some close ties with Reinforcement Learning.[5] It is built upon a bottom-up approach whereby the engineer iteratively adds control layers to the robot, refining its behavior at each step. Lower-level layers such as collision detection take priority over higher-level operations such as identifying objects and reasoning about their behavior. (Brooks 1986; Brooks 1989; Maes et al. 1990; Maes 1994; Brooks 2002)

As an efficient, bottom-up approach to robotics, Nouvelle AI had an important influence on ALife robotic art in the 1990s. Artists such as Louis-Philippe Demers, Bill Vorn, Ken Rinaldo and Simon Penny claim Brooks as a direct inspiration for their work (Rinaldo 1998; Demers and Vorn

[4]According to Whitelaw, Karl Sims (see section 3.2.1) was one of the first artists to relate his practice to the field of Artificial Life (p. 25).

[5]It is important to make a clear distinction between connectionism and New AI. Although conectionism in the field of AI is in direct opposition to the symbolist nature of GOFAI, it shares the same objective of resolving real-life problems in a disembodied, algorithmic way. As such, it is still part of the AI tradition that Brooks aims to challenge through his reclamation of the importance of the body in relation with the real world.

1995; Penny 1997). For example, Penny discusses in a 1997 paper how the Nouvelle AI paradigm contributed to the development of his work *Petit Mal*:

> My project owes a great deal, of course, to Brooks iconoclastic proposals of the 80's such as 'the world is the map'. My goal has been to focus on the social and cultural aspects of the question 'how much can be left out' by concentrating on the dynamics of projection and representation (I mean this latter in a visual and critical theory sense). The tool for this exploration was Petit Mal, an autonomous Robotic Artwork. Petit Mal constitutes an Embodied Cultural Agent: an agent whose function is self reflexive, to engage the public in a consideration of the nature of agency itself. (Penny 1997)

Over the course of the 1990s, Rodney Brooks and his colleagues would develop several robotic projects using these kind of embodied interactive architectures, both with MIT and with the company iRobot, which Brooks founded in 1990 with Colin Angle and Helen Greiner; the same company that released the robotic vacuum cleaner Roomba a decade later. In the late 1990s, Cynthia Breazeal, who had helped on the development of humanoid robot project *Cog* together with Brooks at the MIT's Humanoid Robotics Group, created a robot called *Kismet* that was able to detect and display emotional states. Brazeal has since founded the Personal Robots Group at the MIT Media Lab, where she studies social and emotional robots, such as the robotic pet Leonardo or the smart driving assistant AIDA.[6]

5.4 Enactivism

Nouvelle AI and the aesthetics of behavior have many points of similarity with the concept of *enaction*, first articulated by Francisco J. Varela, Evan Thompson and Eleanor Rosch in their landmark work *The Embodied Mind*. The book offers a critique of computationalism — the dominant approach in cognitive science, which they refer to as *cognitivism* — by bridging Eastern and Western philosophies. Plagued by the image of a pre-existing world whose features are represented by cognitive systems, the authors demonstrate the failure of cognitivism to account for the bidirectional

[6]These systems are known to rely on Machine Learning algorithms, but not so much on connectionist models or reinforcement learning.

nature of interactions between action and perception from the embodied perspective of an agent evolving in the real world (Varela, Thompson, and Rosch 1991, 8).

As explained earlier, connectionism contrasts with the local and symbolic representation model of cognition proposed by the computationalists with distributed and subsymbolic representations learned by self-organizing systems such as neural networks. While the authors recognize the contribution of connectionism to the field, they argue that distributed and numerical representations are still representations. Connectionism thus repeats the same mistake as cognitivism, failing to grasp the importance of the situated body in cognition (9).

The authors thus propose *enactivism* as an alternative to both viewpoints. Based on Mahayana Buddhism, Merleau-Ponty's phenomenology of perception, and autopoiesis theory, this approach suggests to replace the centrality of representation in cognizing agents by recognizing their active involvement in the construction of meaning through autonomous coupling with their environment. In this view, cognition should not be seen as the mere "representation of a pregiven world by a pregiven mind" but rather as the "enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs" (9).

Philosopher and neuroscientist Christine A. Skarda believes that the enactivists' critique of connectionism partly misses the target by conflating two distinct types of connectionist structures: parallel distributed processing (PDP) and self-organized systems (Skarda 1992). While she agrees that the former, exemplified by feedforward pattern recognition neural networks such as the Multi-Layer Perceptron (MLP), suffers from the same representationalism curse as cognitivism, the latter emergentist systems are compatible with enactivism:

> It is misleading to identify, as Varela does, connectionism with self-organizing, *emergentist* systems, and to say that *all* connectionist systems are still wedded to the representations of traditional cognitivism. Some connectionist models are self-organizing, but others are not. All connectionist systems use distributed, highly parallel processing, but that is not the same thing as being self-organizing. PDP systems are susceptible to Varela's attack on representations, self-organized systems are not. I believe that Varela's distinction between emergent and enactive systems is ultimately intended to capture the same fundamental distinction, but it is mistaken to equate emergent systems with connectionism as a whole and set all connectionist systems against the enactive approach.

> This dichotomy is a false one. (266)

Artificial neural networks have been traditionally associated with PDP systems. They are mostly used in *supervised learning* applications, where they are trained to classify or give an estimate, usually based on data tagged by a human "teacher". In other words, they basically learn to give the closest answer to what is expected from them by a human expert. However, what is less known is that MLPs and other neural-inspired variants can be used in a number of *unsupervised* and self-organized ways, such as the Self-Organizing Map (SOM).

Furthermore, of neural networks can be used as function approximators in *reinforcement learning* applications involving real-life agents, where the agent is not trying to recognize specific patterns chosen by human "supervisor", but rather, uses the adaptive qualities that it collects through its connectionist architecture as part of the embodied action-taking process.[7] Therefore, while Skarda stands with the enactivist critique of the PDP branch of connectionism, she also believes that the self-organizing properties of some connectionist systems are "a step in the direction of defining a nonrepresentational alternative in cognitive science" (267).

The vision of an embodied mind that enacts a world of meaning by autonomously coupling with its environment both sustains, criticizes, and feeds upon theories of adaptive systems from the Cybernetics era and 1980s connectionist learning research alike. Indeed, its sheer rejection of the traditional view of enactivism sheerly rejecting connectionism is overblown, and merits refinement. What enactivisists seem to be more accurately in opposition to are certain forms of representationalist connectionism — on the other hand, nonrepresentational neural architectures that rely on self-organization do not seem incompatible with enactivism.

5.5 Coupling

Two concepts lay at the heart of enactivist theory: coupling and autonomy.

> Enaction has been developed over the years as an alternative view to cognition based on minds as abstract symbolic systems, whose fundamental constitutive mode is that of

[7] This is the approach that I took in programming the RL systems described in this research, such as the *Fifth Absence* (2011), *n-Polytope* (2012), and *Plasmosis* (2013).

> a mental representation as a semantic-like correspondence with the world. In contrast, enaction is based on situated, embodied agents, whose world of significance emerges along their active living (?) [sic], not as a representation system, but as constrained imagination, (which the name enaction evokes). More precisely its core theses are twofold: (a) On the one hand, the ongoing coupling of the cognitive agent, a permanent coping that is fundamentally an active embracing of the world in order to in-form it with sense, not a passive reception of it [. . .] (b) On the other hand, the autonomous nature of the cognitive agent understood as an self-produced identity providing a concern (?) [sic] or perspective, an ongoing endogenous activity that it configures into meaningful world items in an unceasing flow. This identity is at the same time natural, since it is based on endogenous configurations (or self-organizing patterns) of complex bodily/ neural activity, yet is also in direct line to subjectivity (Cohen and Varela 2000, . . .)

These two aspects of enactive behavior are complementary to one another. On the one hand, the agent continuously maintains and regenerates itself, keeping its own structure stable while resisting outside perturbations. This *autonomy* is what allows it to define itself as a separate unit from its environment, and to adapt to environmental changes. At the same time, the organism also depends on its milieu to survive. It needs to maintain its *coupling* with the environment since it is precisely in relation to this very environment that it emerges as an embodied entity. "In defining what it is as unity", argues Varela, "in the very same movement it defines what remains exterior to it, that is to say, its surrounding environment." (Varela 1992, 7)

Coupling is an important concept in phenomenology and embodied interaction. It refers to the way by which an object becomes an extension of the human body, not unlike a stick that aids one in moving forward in the dark (Dourish 2001). It is similar to Heidegger's concept of an object being "ready-to-hand". For Heidegger, this happens when the object somehow "disappears" into the background when one is using it. The philosopher gives the example of a hammer. As one interacts with it, as one uses it, it becomes an extension of one's body and one ceases to notice it. However, if one needs to find a way to use the hammer differently, it suddenly "reappears" in the foreground. It is separated from one's body again and one looks at it with a completely different attitude, which Heidegger calls "present-at-hand" (Heidegger 1972).

A distinctive characteristic of adaptive agents lies precisely in their ability to adapt so tightly to objects in their environment that they become "ready-to-hand". In that sense, adaptiveness is a necessary condition to coupling. A compelling consequence, if one considers the case of an adaptive agent's environment as populated by other adaptive agents, is that agents can also become coupled to one another as they adjust to each other's presence. To this end, an adaptive agent is not exactly exactly akin to a hammer or a stick: as one "uses" it until it "disappears in the background", one is also being "used" by the agent: both become "ready-to-each-other's-hand", so to speak. This is especially interesting in the context of interactive interactions, as it shows the potential for nonhuman adaptive systems to generate a different range of aesthetic experiences through bidirectional coupling with human actors.

5.6 Autonomy

Autonomy seems to be a recurring property of intelligent and lifelike systems. In the introductory article to the *Proceedings of the First European Conference on Artificial Life*, Francisco J. Varela and cognitive scientist and economist Paul Bourgine describe the autonomy of agents as the fundamental principle behind Cybernetics and ALife "creatures". "Autonomy in this context", they claim, "refers to their basic and fundamental capacity to be, to assert their existence and to bring forth a world that is significant and pertinent" (Bourgine and Varela 1994, xi), a statement which echoes Varela's enactivist theory. Based on this premise, the authors argue that ALife should move away from trying to understand life by synthetizing its behaviors in software, and rather should engage more seriously with lifelike processes that "assure the key features of autonomy" (xi).

Margaret Boden, an eminent cognitive scientist at the University of Sussex, delineates three different aspects of autonomy in living and artificial systems. First, she claims that autonomy implies an indirect response to the environment, where reactions are modulated by experience. Second, autonomy supposes a self-generated control mechanism, as opposed to one that would be pre-given or scripted — such as emergent behaviors that appear in artworks involving SOMs such as Yves Amu Klein's *Octofungi* (Klein 1998; Klein and Hudson 2003) or Nicolas Baginsky's *The Three Sirens* (1992—2005). The third component is defined by Boden as "the extent to which inner directing mechanisms can be reflected upon, and/or selectively modified" to adapt to the current context (Boden 1996, 104).

The first and third facets of autonomy seem directly connected to adaptation and learning: simple experience-driven reactions in one case, and the more complex ability to analyze one's own behavior in the other. Yet the second aspect described by Boden concerns the emergent properties of the system, which also happens to be a crucially important idea in the study of adaptive artificial agents.

5.7 Emergence

Emergence refers to the mechanism whereby higher-order forms or processes emanate from the complex interactions of lower-order units. Emergence has been widely studied by scholars interested in questions of artificial cognition and living systems. It is often associated with self-organization, such as in the work of ALife researchers, cyberneticians and connectionists. However, emergence also evokes an idea that goes beyond the automated configuration of a system: the generation of novelty (Soler-Adillon 2015).

Peter A. Cariani is an interdisciplinary researcher who has developed one of the most compelling theoretical model on the role of adaptation in cybernetic and ALife systems through an original and constructive critique of computationalism (Cariani 1989; Cariani 1990, 1991). He has contributed a uniquely stimulating taxonomy of artificial systems that establishes a clear relationship between adaptation and emergence. Through the precision of his theorectical framework, he distances himself from phenomenological critiques of computationalism such as Dreyfus' *What Computers Can't Do* (Dreyfus 1979) or Searle's "Chinese room" (Searle 1980), and rather situates himself in the tradition of American pragmatists such as William James and John Dewey.

Cariani differentiates cybernetic devices on the basis of their adaptive qualities, identifying three kinds of such systems: formal, adaptive and evolutionary. *Formal* devices are purely (formal-computational) or partly (formal-robotic) symbolic apparatuses that respond to a fixed set of instructions: thus, they are nonadaptive. *Adaptive* systems are capable of adapting their computational structure based on experience, but are limited by their fixed semantical components (sensors and effectors). Machine Learning models such as MLPs, GAs, and even adaptive robotic agents are part of that category.

The third category, which Cariani calls *evolutionary* devices, are able to adaptively construct their own sets of sensors and effectors based on their interactions with the environment. Such devices thus consist of "a set of sensors, a set of effectors, a computational part, a performance measure, and an apparatus for constructing new sensors and effectors" (Cariani 1989, 132).

This category can be refined by considering systems that have adaptive semantics but a non-adaptive syntactic part, such as the immune system. *General evolutionary* devices are those that are both *adaptive* and *evolutionary*: in other words, general evolutionary devices that display both semantic and syntactic adaptiveness, and there are plenty of examples of such systems in the biological world. However, there exists to the author's knowledge only one example of such a human-built system: Gordon Pask's electrochemical adaptive assemblage which allowed the evolution of a primitive "ear" (27).

The main "advantage" of evolutionary devices as compared to adaptive or formal systems lies in their open-endedness, in other words, their ability to generate novelty, which Cariani directly associates with the question of emergence.

> The problem of emergence is useful in evaluating the open or closed nature of the devices in our taxonomy precisely because it relates to the problem of novelty in the world. If we want to enlarge our own capabilities and free ourselves of the burden of complete specification, our devices must be creative. If we want our devices to be creative in any meaningful sense of the word, they must be capable of emergent behavior, of implementing functions we have not specified. Our emergent devices must not be prisoners of our notational systems if they are to aid us in our own break-out. (148)

Device type	Plasticity	Capacities	Limitations
Formal-computational	Fixed syntax	Reliable execution of pre-specified rules	Limited to pre-specified rules and states
Formal-robotic	Fixed syntax Fixed semantics	Reliable execution of fixed percept-action combinations	No feedback or learning from environment
Adaptive	Adaptive syntax Fixed semantics	Performance-dependent optimization of percept-action coordination	Limited to percept and action categories fixed by the sensors and effectors
General evolutionary	Adaptive syntax Adaptive semantics	Creation of new percept and action categories; Performance-dependent optimization within these categories	Time to construct and test new sensors/effectors may be very long

Table 3: Summary of Cariani's taxonomy of devices.[8]

Three major theories of emergence are examined by Cariani: computational emergence, thermodynamic emergence, and emergence relative to a model. *Computational emergence* stands for the computationalist theory of self-organizing systems, similar in viewpoints to those held by proponents of "strong ALife" like Langton and Ray. It assumes that all emergent behavior at the macro level is reducible to micro level rules. Furthermore, proponents of this viewpoint argue that emergent structures such as living systems can *actually* be realized by such symbolic operations *themselves*, not just in the mind of the observer. *Thermodynamic emergence* theories attempt to describe emergence employing differential equations such as those used in physics. Contrary to computationalists, proponents of thermodynamic emergence do not suggest that the computation of these equations should be considered as actualizations of the emergent systems they describe.

Emergence relative to a model (or "observer-centric emergence") was first developed by theoretical biologist Robert Rosen and defines an emergent event as "a deviation of the behavior of the physical system under observation from its predicted behavior" (30). In other words, emergence comes from the fact that since we dispose of only a finite number of observable dimensions, in a universe which contains a potentially infinite number of attributes, it follows that our models of the world are always incomplete accounts of it. (157)

> Emergence relative to a model, then is a result of the finite and hence incomplete character of all models of the world. At some point in time we can, if we are fortunate,

> construct a model which will deterministically capture the behavior of the physical system. The behavior predicted by the model will, for some period of time, correspond to the observed behavior of the physical system, because it was constructed to do so. But eventually, if one waits long enough, all physical systems will diverge from their models, but some will diverge before others. (157)

Emergence relative to a model allows an integration of adaptation and emergence in a comprehensive framework. The taxonomy of adaptivity at the core of Cariani's theory can now be attached to the emergent qualities of a system's behavior:

> When the behavior of the physical system, in this case the device itself, bifurcates from the behavior of the model, another model will have to be constructed which will capture subsequent behavior of the physical system/device. [. . .] We will call the situation where changing the computational part of the model is sufficient to recapture the behavior of the physical device *syntactic-emergence*. Only new syntactic linkages need be formed. We will call the situation where adding new observables is necessary to recapture the behavior of the device *semantic-emergence*. These types of emergence correspond to device types in the adaptivity taxonomy. *Formal devices are nonemergent. Adaptive devices have syntactic-emergent behavior. Evolutionary devices have semantic-emergent behavior.* (158)

This "bifurcation" from the model's behavior is thus, according to Cariani and Rosen, the locus of novelty emergence in the agent's behavior. Emergence is realized by the agent through its adaptive capabilities, either syntactic, semantic, or both. As such, one could say that adaptivity is the means by which emergence is realized in adaptive and evolutionary systems. In that context, adaptivity is seen not just as a way for agents to self-organize, but as a necessary condition for creativity.

Cariani's viewpoint on the central role of adaptation in the emergence of behavior of agent-based systems is particularly enlightening. It has an important consequence in terms of aesthetics. Emergence is a necessary condition for adaptation in emergent-relative-to-a-model systems, because adaptation is precisely what steers the self-organization of such agents. Yet, emergence is not a sufficient condition for adaptation. Indeed, most emergent systems found in nature are nonadaptive. For example, consider thermodynamic complex systems such as meteorological phenomena (e.g.,

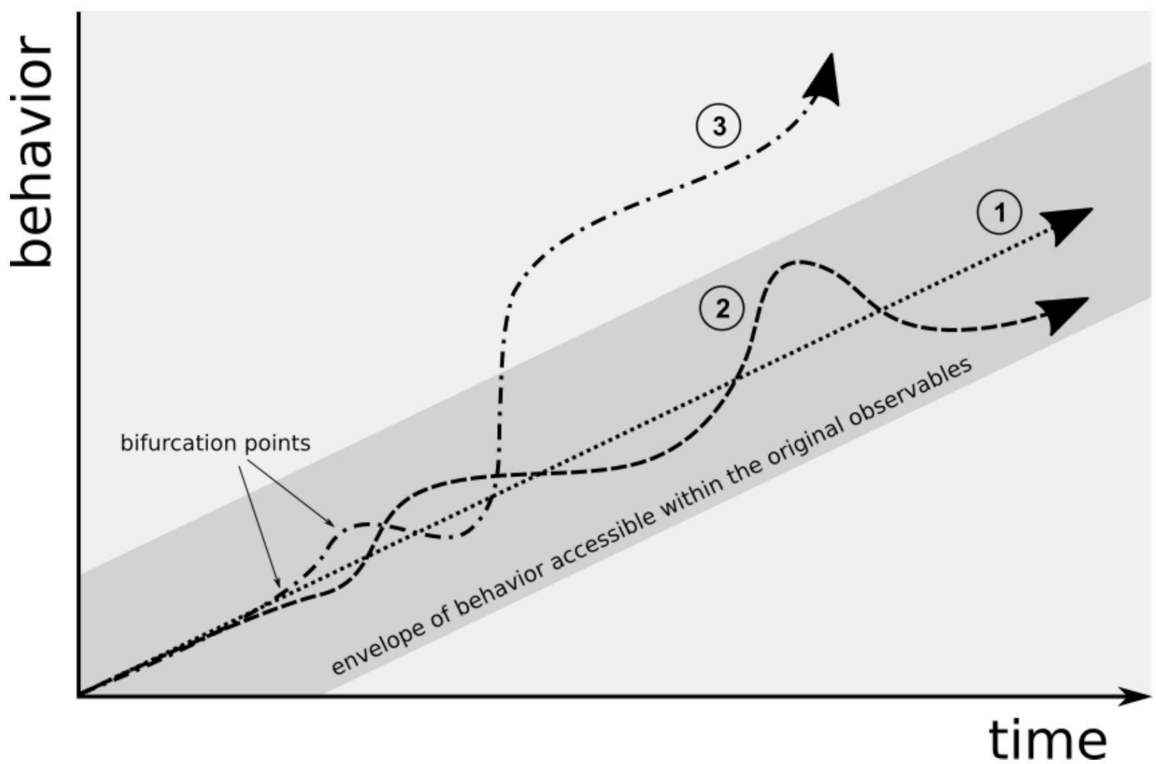


Figure 31: “Divergence of the behaviors of adaptive devices from fixed models of them”. (1) Formal-computational and formal-robotic (nonemergent); (2) Adaptive device (syntactic emergent); (3) Evolutionary device (semantic emergent). Adapted from (Cariani 1989, 31).

cloud formation, precipitations, hurricanes, etc.) and gravitational systems such as solar systems and galaxies, or computational simulations of such phenomena (e.g., particle systems in physics engines).

Now, following a similar set of intuitions, one can infer that adaptation is a necessary, yet insufficient, condition of life. On the one hand, adaptation seems to be a defining feature of life (this is one of the central claims of Cybernetics and Artificial Life), appearing first and foremost through evolution, and, in the most advanced systems, in real-time.[9] On the other hand, there exist adaptive systems whose status as living beings is at the very least frought, such as Cybernetics devices like Ashby's *homeostat* and Pask's artificial "ear".

Hence, while adaptation is not enough to define life, it lies "one step closer" to it than emergence. That is not to say that emergence has a less important contribution than adaptation to the lifelikeness of an artificial behavior — intuitively, I would be inclined to think that it contributes to most of it. Instead, emergence implies that self-organizing behaviors without adaptive properties might lack some of the affective components possessed by living entities, which would give a more substantial impression of being in the presence of one. This is significant for this study, as it points towards the importance of adaptation in the building of lifelike artistic systems.

5.8 Authorship

Digital artist Marc Downie, who worked early on several interactive pieces with artificial social characters such as *alphaWolf* (Tomlinson and Blumberg 2002) and *The Music Creatures* (2000—2003), brings an interesting perspective to the question of agent-based behaviors. He criticizes two of the most common concepts in the field of interactive art: mapping and emergence.

Mapping is a very common metaphor in media art. It represents a transformation of one signal into another, which Downie finds extremely limiting. He suggests, as a replacement, the concept of an *agent*, which has the capacity to be embedded into its world.

However, Downie explains that emergence is also problematic in artistic creation when contrasted

[9] As brought up above, Margaret Boden mentions how real autonomy in agential systems requires adaptation (Boden 1996).

with the question of *authorship*. He argues that emergence-based approaches try to avoid the question of authorship altogether by trying to create processes that work by themselves, without human intervention (Downie 2005, 29). Yet, despite decades of efforts, we are still waiting for the advent of higher-order emergent artificial life structures (Bedau 2000).

> Some of the motivation for the agent-based — and other distinct but related trends in the 80s and 90s such as connectionism and artificial life — came from an often open and explicit authorship twist: that reactive, connective, adaptive or behavior-based systems avoid the burden of knowledge engineering (i.e. knowledge authorship) and exploit a far closer relationship with statistical machine-learning techniques to avoid the hand-tuning, assembly or even creation of systems altogether. (Downie 2005, 29)

Moreover, as Downie claims, the main difficulty faced by digital artists is not so much in "generating potential", because this is the part made relatively easy by computer technology. The trouble is rather to put that potential to work in the creation of a piece of art. Therefore, "there is no need to be excited should it turn up or rather emerge without much effort on our part." (36) According to this logic, we should instead focus on hybrid systems that make integrate adaptive systems with more traditional AI such as rule-based and goal-drected components.

In other words, once the artist's fascination for emergence drops, what really matters is the aesthetic experience of the work, which, for Downie, is sourced from the artist who authored the work through the creative process. The question is not so much how autonomy, adaptation and open-endedness affect our relationship with cultural agents *per se*. Rather, it concerns the kinds of effects that can be created by integrating them in a creative practice of artificial agents in an effort to provoke evocative relational experiences.

So when thinking about such adaptive agents, what an artistic researcher should focus on is the sociocultural context in which the artist intervenes in relation with the data that is available. By nature, adaptive agents come with distinctive authorship questions. What are the possible actions? What data will be fed into the system? With what criteria are the agents trained (i.e., what is the evaluation function)? How do the adaptive features of the agent help (or hinder) what is meant to be expressed by the system? In particular, the adaptive nature could lead to tighter autonomy, and may give a sense of precariousness and uniqueness to the work: which are strong features that can be exploited by artists in the production of aesthetic experiences.

5.9 Believability

So what exactly does the authoring adaptive agents entail? What are the characteristics that one should look for when designing such agents? Downie points to the research on interactive drama carried in the 1990s at Carnegie Mellon under the scope of the *Oz Project* as offering potential leads. The research program, led by computer scientist Joseph Bates, tried to develop technologies that would allow artists to design complex dramatic interactive fictions. In his famous paper "The Role of Emotions in Believable Agents", Bates suggests that researchers in AI are wrong in trying to create machines that act like humans trying to reproduce thinking, reasoning, and learning. Instead, he proposes they should follow in the footsteps of artists who rather attempt to make their characters *believable* by having them display recognizable emotional states (Bates 1994).

Bates describes how he combined a goal-directed, representation-free architecture inspired by Nouvelle AI with an emotion generation architecture to implement an ensemble of "believable" characters called "Woggles" in a 1992 piece called *Edge of Intention*. Recalling the work of Disney animators Thomas and Johnston, Bates claims that what makes us care about artificial characters lies in what we recognize as their emotions and desires. "If the character does not react emotionally to events, if they don't care, then neither will we. The emotionless character is lifeless, as a machine." (123)

By turning the discussion on intelligent agents towards emotions, Bates echoes an interest in Affective Computing, a field that flowered in the mid–1990s. The mother of this field, Rosalind Picard at the MIT Media Lab, claims that emotions are an essential part of human intelligence and should thus be considered at least on equal grounds with rational and abstraction capabilities by AI scientists (Picard 2000). Social robotics expert Cynthia Breazeal considers that autonomy alone is not "sufficiently life-like" and argues that believability is an important aspect in the design of social robotic agents because it projects the "illusion of life" and gives the agent a personality (Breazeal 2002, 8).

Following Bates, Michael Mateas adopts a more general point of view by associating the focus on authorship adopted by artists such as Downies with a novel branch of research he calls *Expressive AI* that lies between traditional AI (GOFAI) and Nouvelle AI (which he refects to as *interactionist AI*). Expressive AI practictioners are created cultural artifacts that behave in a seemingly intelligent way within a specific sociocultural situation (Mateas 2001). In such a setting, the system "expresses the author's ideas within a performative space and is both a messenger for and a message from the author." (150)

This suggests that Machine Learning systems such as neural networks might be giving us a false impression inherited from GOFAI, that impression of being neutral structures that can learn almost anything. Sure, these systems are powerful models, able to perceive correlations out of high-dimensional data streams; but their performance is nonetheless extremely influenced by the kind of data that flows into them. In other words, choosing different inputs and outputs in these systems comprises an important editorial decision, as it will directly modify the resulting predictions of the neural net, hence the behavior of the agent in the reinforcement learning systems.

5.10 Social Agents

Science and technology theorist Sherry Turkle has studied the social dimension of computational agents and robotic systems throughout her career. In particular, she studied the effect of interactive toys such as Furby, AIBO, Pleo, and "My Real Baby"; a class of technological devices that are designed for social interaction. Turkle refers to these toys as "relational artifacts", and suggests that the traditional view in computer science, which posits that computational tools such as these items "do things *for* us", is flawed. These agents rather "do things *to* us" by changing the way we perceive ourselves and our sociotechnical environment. Their dissemination in households might even be fostering a new "robotics culture". (Turkle 2006, 1)

These agents, Turkle claims, act in two ways. First, affectively, by being things on which we simply project our emotions—a phenomenon which Turkle refers to as the "Rorschach effect". This is the space occupied by traditional computational objects such as the artificial chatterbot

ELIZA (Weizenbaum 1976). Second, Turkle claims that contemporary relational objects such as AIBO and Pleo allow a move from this projection-based, individualistic perspective to that of a "psychology of engagement", an "evocative object effect" that is cognitive more than affective (Turkle 2006, 2). These two effects — projectional/affective and evocative/cognitive — are not incompatible nor completely independent, but are perhaps best described as strongly interrelated in human-robot relationships (4).

Turkle notices that autonomy (c.f., section 5.6) in artificial agents is often associated with *aliveness* (the property of being alive) as if these agents were "alive in a way" (Turkle 2006). This belief, and the way it is expressed, directly contrasts with people's relationship with computers or other technological devices that are not inherently social nor autonomous in of themselves. Describing a change from people's (especially children's) beliefs about computational artifacts in the early 1980s and robotic creatures of the mid–1990s onwards, she writes:

> With relational artifacts, the locus of discussion about whether computational artifacts might be alive moved from the psychology of projection to the psychology of engagement, from Rorschach to relationship, from creature competency to creature connection. Children and seniors already talk about an "animal kind of alive" and a "Furby kind of alive." The question ahead is whether they will also come to talk about a "people kind of love" and a "robot kind of love." (8)

Can adaptation in artificial agents contribute further to their aliveness, and thus to their propensity for evocation? The learning capabilities of AIBO and Furby seem to be key features of their commercial success, and also seem to add value to their long-term appreciation. Being patented commercial products, it is difficult to access the precise methods for how their softwares are implemented. However, first-hand experience with these toys suggests that they rely to a very large extent on a relatively closed form of adaptation that is simulated rather than enacted: in other words, they all seem to be pre-tuned for learning certain specific things through their interaction with humans. This can reduce their value over time once their owner has explored all of the possibilities offered by these devices.

In other words, these electronic "pets" lack the kind of indeterminate and open-ended qualities that seem to be present in many of the works considered in this thesis which allow for a wide range of social interactions to happen, rather than being constrained to a limited number of possibilities. How does this open-endedness affect social interactions with the robotic agencies in question? Addressing this question adequately would require experimenting with agents embedded in everyday life for long periods of time, which is beyond the scope of this thesis. However, an indeterminate and complex nature of Machine Learning agents, the particular goals these agents would aim to achieve, and the way that agent behavior unfolds through time are all important items for consideration when studying their social impact.

5.11 Performativity

One last concept worth examining is that of performativity, which originated in the 1950s through the work of language philosopher J. L. Austin. Defying analytic philosophy, Austin argues in *How to Do Things With Words* that most assertions made in discourse are not statements describing a reality, but are rather *doing* something in the world. Austin introduces iterability as a necessary condition of a successful “performative utterance”: the citation or reiteration of a sentence — such as the “I do” sealing a marriage — is what makes it efficient as a “speech act” (Austin 1962).

There are parallels to be drawn between adaptive behaviors and iterability. In response to environmental changes, adaptive agents such as those who use Reinforcement Learning (see section 3.2.1) tend to develop and repeat strategies it believes are best fit to the situation, often in the shape of recognizable patterns of actions. From time to time, it explores new actions and if they appear to have good results, it will tend to promote these new actions in the future: otherwise, the agent will strengthen its current strategy.

The agent thus works smoothly and continuously to the articulation and re-articulation of contingencies, patterns and regularities of its world, through its behavioral model. Each action has an influence on its world as it changes both its environment and its internal structure (and logically, its future behavior).

At each step, the agent performs a more or less approximate citation of a single action or sub-sequence of actions performed in the past. These past actions are also based on previous actions and so on. In short, every action can be seen as a modification of a pre-existing citation, of a "script" that is constantly actualized. We can therefore say that such an action is repeatable, readable, contextualizable.

Performativity has become, over the years, an attractor for a vast body of research in various disciplines, such as performance studies (Turner 1982; Fischer-Lichte 2008; Schechner 2003), gender studies (Sedgwick 1993; Butler 1999, 2004), and economics (Callon 2006). One of its great strenghts is to provide an alternative to representative accounts of science, society, and cognition. Andrew Pickering describes this "performative move" in his analysis of human relations with its environment. Discussing his theory first put forward in his book *The Mangle of Practice*, he explains that the "key move" was for him to "focus on performance rather than cognition".

> We have all been taught to think of science as primarily a cognitive activity – the production of knowledge about the world – but my argument was that if you want to understand scientific practice, you should start by thinking about (a) the performance of scientists – what scientists do; (b) the performance of the material world – what things do in the lab; and (c) how those performances are interlaced with one another.

He relates his conception of performativity to that of agency, a term which to him refers "directly to action, doing things that are consequential in the world". He uses these notions to describe the interactions between humans and nonhumans in scientific practice as a performative, embodied interaction: a "dance of agency" (Pickering 2013, 1–2).

Of particular interest to this study is Pickering's more recent application of this theory to the case of Cybernetics, a "postwar science of the adaptive brain" (Pickering 2010, 6). Cyberneticians, he claims, did not see the brain as an apparatus able to generate representations and manipulate them in an orderly manner — as would be the case for most of their successors — but rather as an active organ that does things in the world. In summary, Pickering claims, "the cybernetic brain was not representational but performative" and "its role in performance was adaptation." (6)

Pickering highlights the peculiar character of adaptive systems as introduced by cyberneticians:

> There is something strange and striking about adaptive mechanisms. Most of the examples of engineering that come to mind are not adaptive. Bridges and buildings, lathes and power presses, cars, televisions, computers, are all designed to be indifferent to their

> environment, to withstand fluctuations, not to adapt to them. The best bridge is one that just stands there, whatever the weather. Cybernetic devices, in contrast, explicitly aimed to be sensitive and responsive to changes in the world around them, and this endowed them with a disconcerting, quasi-magical, disturbingly lifelike quality." (7)

What is especially compelling about Pickering's analysis is that what he says about scientific practice can also largely be said about the art practice of adaptive embodied agents, which deploys technical creative methodologies similar to those used by cyberneticians such as Ashby, Pask and Walter. Pickering's posthumanist perspective harbors a non-anthropocentric vision over agency, which suggests that in the context of an agent-based installation, both the artist, the recipients, and the artificial agents are all considered active participants in the unfolding of a performative aesthetic experience. This suggests an alternative perspective over artmaking, and the relationship between art and science, which I will further discuss in chapter 6.

5.12 Behavior Morphologies

Now that the notion of adaptation has been more precisely defined, it is important, before moving forward, to describe in more detail what is meant by "embodied" or "situated" in regards to agents in an artistic context. The aesthetic framework that I want to articulate here resonates with the work of Simon Penny. Directly inspired by Rodney Brooks' revolutionary work on situated robotics from the late 1980s that critiques representational systems in AI, Penny argues for a new "aesthetics of behavior" that contains a rejection of computationalism:

> I felt that underlying the fundamental premises of computer technology is the acceptance of Cartesian dualism, the separation of the mind and body. This separation is written right into the technology as hardware and software. It is inscribed into the fundamental premises of computer science.

He further explains:

> Part of my project has been to try to find theoretical resources to build a new aesthetics around a rejection of these premises to formulate what I refer to as an 'aesthetics of behavior'. It is premised on the idea that when we use real time computational

> technologies for cultural practice we are doing a new aesthetic practice, which involves the designing of behavior. We are somehow building a contingent model for what might happen in the world, and how our system might respond in order to direct the aesthetic attention of the user to a direction consistent with the artwork itself. It is a complex and new aesthetic negotiation of the dynamics of interaction and authorial intent. (Kim and Galvin 2012, 138)

Penny hereby joins Brooks and Dreyfus in their critique of the dualistic vision of behavior and cognition that taints classical AI. Behavior, he claims, should not be understood as a purely computational, disembodied thing called “software”, but rather needs to be grasped as a situated process running through an agent’s body. Or course, behavior in a computational-based artwork has algorithmic components, however, in the hands of the artist the code becomes another material, with its own specific characteristics to be integrated with visual, sonic and physical components in the construction of a global aesthetic experience.[10]

Whereas both Brooks’ *Nouvelle AI* and Penny’s behavior aesthetics are characterized by their reliance on a “bottom-up” approach to technical practice directly inspired from ALife research, their strong anti-computationalist stance is also directed at the rampant computationalism characteristic of 1980s ALife. Thus, both Nouvelle AI and behavior aesthetics rearticulate concepts of emergence and self-organization in ALife by integrating them in a performative theory of behavior that places the agent’s body at the center of the equation. As such, Penny’s proposed artistic framework is constitutionally different from concurrent disembodied artforms such as Algorithmic Art — that essentially aim to produce stabilized forms, usually computer-generated images — and an important part of Artificial Life Art that generates time-based simulations on the computer.

As an artist working with agent-based systems, I concur with the anti-computationalists: life and cognition are not “pure” processes that can be separated from a sensorimotor body running in the physical world. I hereby align with Harnad’s claim that cognition is at least *partly* non-computational, though *some* computation (i.e., rule-based symbol manipulation) might be involved in it (Harnad 2001, 2008). I join my voice with that of Simon Penny, arguing for a new field of aesthetics opened up by computer technologies, with *behavior* as its central concept. Yet, I believe

[10]Questions regarding the material practice of programming artificial agents for media installation is further discussed in chapters 4 and 6.

there are still important missing pieces in our understanding of the actual aesthetic qualities of such behaviors.

Gordon Pask's own definition of behaviors, which he detailed in his 1968 book on Cybernetics, offers a visionary perspective over behaviors that connect well with Penny, while still allowing for a formalization in terms of their morphological evolution. In line with his view, I argue that behaviors are best defined not as algorithmic recipes, but rather as real-time material patterns as they are recognized by an observing entity. As Pask writes:

> As observers we expect the environment to change and try to describe those features that remain unchanged with the passage of time. An unchanging form of events due to the activity within an assembly is called a behavior. (Pask 1968, 18)

There are two important implications of this definition. First, while an agent's behavior involves a sequence of events that constantly change over time, its behavior has a recognizable "shape" that remains temporally invariant. Pask gives the example of a cat, which consists of "performances like eating and sleeping and, once again, it is an invariant form selected from the multitude of things a cat might possibly do" (18).

Second, while a behavior is always generated by a system — which could, but need not be, computational — it only exists through its perceptual effect on an observer. This implication is particularly appropriate to an aesthetic framework, as it focuses on the phenomenological experience generated by the agent-based performance, as it unfolds through time and space in the material world. This connects directly, in fact, to the pragmatic aesthetics of John Dewey, who claims that works of art should not be thought of as objects, but really as "refined and intensified forms of experience" (Dewey 1959, 3).

I posit that different categories of system architectures allow for different kinds of behaviors, thus allowing the emergence of different aesthetic experiences. What interests me here is to further analyze Penny's artistic frame of reference by looking more closely at embodied agents with adaptive qualities. Existing taxonomies of Cybernetics systems have mainly focused on relational and structural aspects of these systems (Rosenblueth, Wiener, and Bigelow 1943; Cariani 1989). In this section, I propose a flexible taxonomy of embodied systems that focuses on the aesthetics of agent behaviors as their shape unfold in time.

The "zero-degree" of that categorization is the "behaviorness" of the system, that is, whether it should be considered to have a behavior or not. The initial differentiation criterion, I argue, lies in the structural capacities of the system, more precisely in the existence of an internal state. Stateless devices are akin to mathematical functions: their outputs/actions only depend on their inputs/observations. By design, they are incapable of accumulating experience.

Such systems are known in the the field of digital media art as mappings. Their widespread popularity is evidenced by the prevalence of data-flow softwares such as Max/MSP or PureData, often appearing under names such as "visualisation" or "sonification". Downie heavily criticizes this hegemony of mapping in interactive arts. He argues that its apparent generality, which is seen as beneficial, makes it ineffective and sterile: precisely because its definition has "no limits" it also has "no use". He writes:

> In practice one can sense in this "function-like" aspect of mapping is a kind of college-level, piecewise linear or otherwise smooth, locally stationary, state-less, typically decomposable relationship between input and output. Such a vision acts as a normative idea of how, in this field, numbers get transformed into numbers. The best work in the field, of course, pushes against this central tendency, but the rules and arena remain fixed. (Downie 2005, 17)

Devoid of any kind of autonomy and agency, mapping-based devices are behaviorless, their conduct relying almost entirely upon the data that is fed into them. Whatever sense of aliveness associated with them truly lies in the system that generates this data, be it a human performer or a natural phenomenon. Their statelessness imprisons their "performance" into the instant: their world, if they have any, is a succession of independent moments. They are, in other words, zero-order behaviors (i.e., "nonbehaviors").[11]

Agent-based systems, which are the focus of both this dissertation as well as Downie's, are behaviorful in their ability to extend their world into the past through the use of some kind of inner structure. These stateful devices possess some sort of "memory" (whether it is discrete, continuous,

[11][Pask's example of the "behaviour of a statue" is an extreme case of such a "nonbehavior" (which he actually chooses to ignore). (Pask 1968, p. 18)

long or short) which is modified by their interactions with the environment. In other words: their past experiences influence their present actions (at least within a certain time window).

This statefulness, which in other words implies some form of structure or trace, can be found in a side variety of computer programs. For instance, formal devices as defined by Cariani can possess states, typically recognizable in computer code as named variables of different types (i.e., booleans, integers, floats), however these syntactic components are fixed. Behaviors generated by these systems are thus bound within a certain domain. Hence, while an agent's response to sensory data may change depending on context, its behavior itself does not change through time. Given enough time, it will, inexorably, come to repeat similar patterns. We will thus refer to these conducts as first-order behaviors.

To understand this idea better, consider how a behavior can have a certain, recognizable morphology that exists in a domain different from other forms of non-computational, "stabilized" media, so to speak, such as an image or video, or even, as I explained earlier, real-time mappings such as sonifications or visualizations. The shape of a behavior is parameterized by the sensors, effectors and processing capacities of the system that generates it, and evolves within a certain space-time territory. Morphology and morphological processes have been used to describe time-based behaviors in the writings of contemporary music composers such as Iannis Xenakis and Agostino Di Scipio (Xenakis 1981a; Di Scipio 1994; Solomos 2006).

Because of their inability to generate new forms and/or to transform their own form, I argue that the behavioral morphologies produced by formal, rule-based systems, are fundamentally different from those produced by adaptive and evolutionary agents. The latter produce second-order behaviors (i.e., "metabehaviors"), which involves the coming-into-being, and possibly transformation, of their own (first-order) behavior. They therefore exist in a "different time" than their formal/fixed counterparts, which affects the overall aesthetic effect they can engender.

I propose to use the concepts of morphogenesis, morphostasis and metamorphosis to further characterize the different processes by which behavioral morphologies exist, emerge and/or change over time. These notions are related, each in their own way, to ideas of emergence, self-organization, self-regulation, novelty and autonomy. As these ideas bring processes related to forms to the fore,

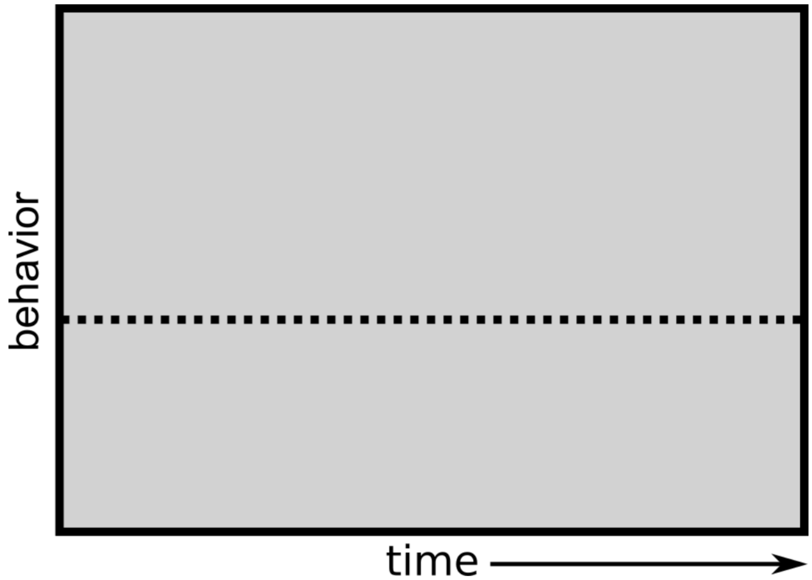


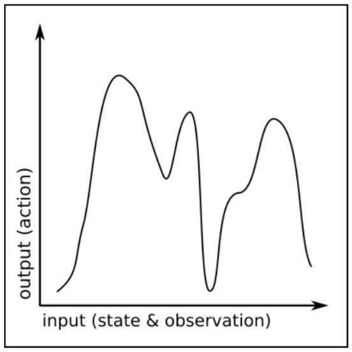
Figure 32: Example temporal evolution of a first-order behavior. The vertical axis represents the *behavior* of the system, understood as the temporally invariant shape of observable events the system generates through its actions[12]. The horizontal axis represents the advance of time. The graphic shows how first-order behaviors remain temporally stable.

they seem particularly appropriate to support an aesthetics of behavior.

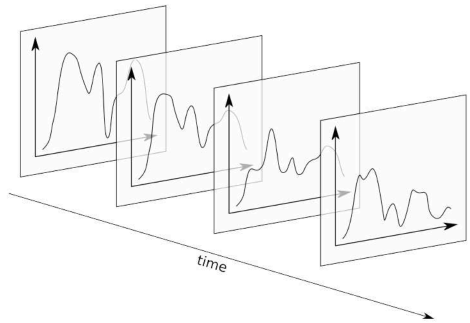
Morphostasis refers to the process whereby a behavior hovers around a stable state of being. While the behavioral patterns might look like they are changing when considered over a certain period of time, morphostatic behaviors quickly exhaust the space of dynamic patterns they can generate and start 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, are able to 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. In essence, the term should be understood quite similarly to the way it is used in common parlance: that is, as an outstanding transformation in



(a) First-order behaviors: Formal systems react to the world in a way that is more or less stable through time.



(b) Second-order behaviors: Adaptive and evolutionary agents modify their behavior in reaction to observations.

Figure 33: Representation of behaviors in nonadaptive vs adaptive agents. This figure compares the response of formal systems with that of adaptive and evolutionary systems. The relationship between inputs (i.e., what the agent observes as well as its current state/memory) and outputs (ie., the actions taken by the agent) of a formal system is temporally invariant, whereas adaptive and evolutionary systems allow it to change through time.

Table 4: Orders of behavior in agent-based systems.

Order	Type	Device types	Properties
0^{th}	Nonbehavior	Stateless functions (mappings)	Actions are purely dependent on observations (no memory).
1^{st}	Behavior	Formal-computational Formal-robotic	Actions depend upon both a set of observations and a state (memory). Behavior is unable to adapt or evolve.
2^{nd}	Metabehavior	Adaptive Evolutionary	The behavior itself transforms through time in response to the environment.

a living being or thing. The two main dimensions of metamorphosis are (1) the *metaboly*, that is, the magnitude of the transformation undergone by the behavior; and (2) the *speed* at which the behavior transits from one form into the other.

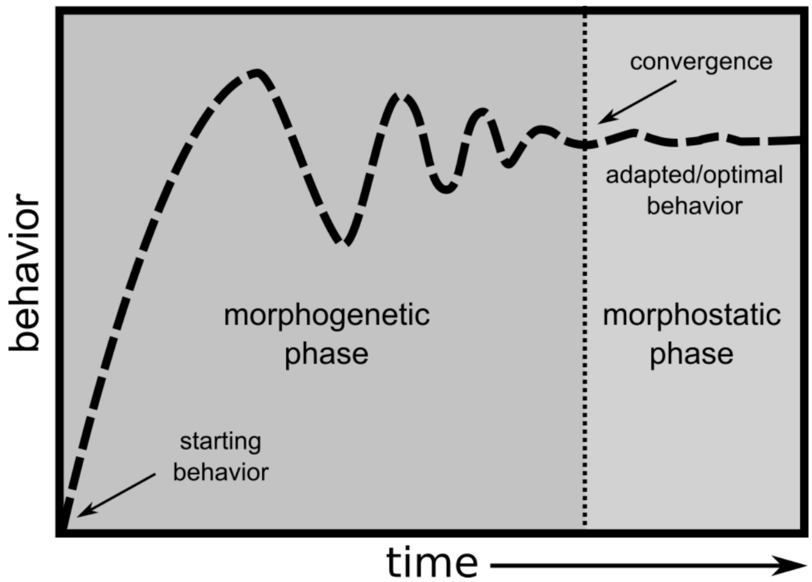


Figure 34: Example temporal evolution of an adaptive behavior. Distance along the vertical axis represent difference in the form of observable events produced by the agent. The graphic shows how second-order, adaptive behaviors iteratively change through time through a process of morphogenesis, until they stabilize into an optimal first-order behavior, thus entering a phase of morphostasis.

These aspects of an agent's performance should be seen less as hard-set categories, but rather as conceptual tools for describing processes of behavior formation. These qualifiers complete concepts such as those previously discussed (i.e., enactivism, coupling, autonomy, and emergence) by bringing attention to the morphology of behaviors and their evolution.

From this perspective, both formal systems as well as self-regulated devices such as the homeostat, or pre-trained Machine Learning algorithms, produce purely morphostatic behaviors. However, they are distinct from each other in the kinds of first-order, repetitive patterns they produce, which are related to their different structural and behavioral properties, as highlighted above.

At the opposite end of the spectrum, some morphogenetic systems freely move from one behavioral embodiment into another, living in a constant state of metamorphose, as if never fully coming into being. These systems are often referred to as "generative": they evolve behaviors regardless of their fitness or value (Bown 2012).

Adaptive systems, on the other hand, evolve their morphologies in relationship to a usually indeterminate "ideal" (i.e., optimal in regards to whatever the evaluation function is) behavior, which they try to approach and match. In this, they differ from *nonadaptive* second-order behaviors. Adaptation, like intentionality, requires an object: systems do not simply adapt, they adapt *to* something. Adaptive systems are relational devices by definition: they are governed by their coupling with another behavior, which in turn can be of zeroth-, first-, or second-order. Their experiences effects their inner structure so as to improve their prospective performances. In other words, their past feeds their future.

Typically starting from a state of pure randomness, adaptive agents run through a learning process of morphogenesis where they progressively and asymptotically modify the shape of their behavior to better perform in relationship to their evaluation function. When they reach their final form, they enter a state of morphostasis, exploiting the stabilized, learned behavior which they converged to. Some adaptive systems have the ability to depart from this crystalized demeanor, either as a result of an internal intentionality, or as a response to environmental changes that require drastic adjustments to their performance.

The aesthetic experience of these behaviors is dependent on a number of factors. The ratio between the magnitude of change and the time period necessary to perform it 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 uncaninness. For example, in *Vessels*, the robots are always in a state of flux, which might explain the feeling of estrangement inspired in some members of the audience. As the audience is never fully able to observe a recognizable behavioral pattern, to some of its members, the robots' behavior seems purely random.

Finally, adaptive behaviors convey a certain narrative. Unfolding before our eyes, we perceive

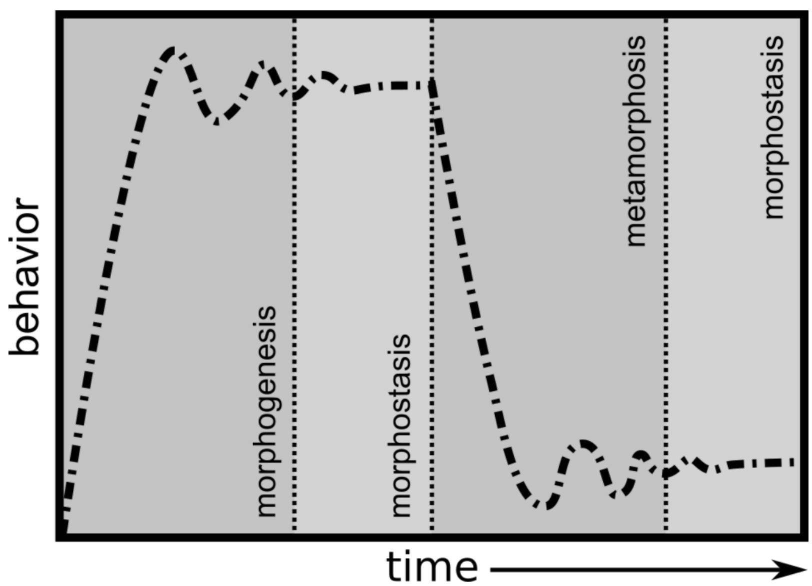


Figure 35: Example of the temporal evolution of an adaptive behavior going through multiple phases of learning. Starting from a random behavior, it runs through a morphogenetic phase until it converges to an optimal behavior, stabilizing into morphostasis. Then, subjected to environmental changes, it needs to readjust itself, metamorphosing into another shape that performs better in the new conditions.

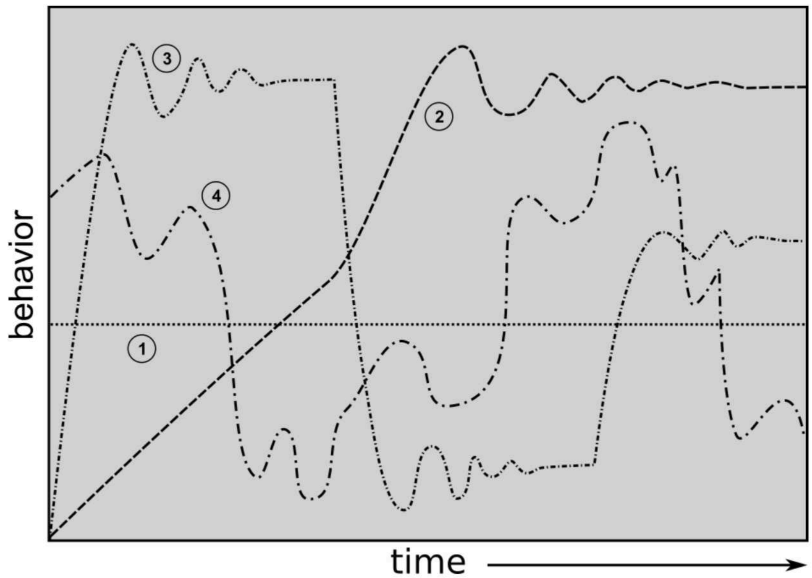


Figure 36: Example of the temporal evolution of different kinds of behaviors: (1) first-order behavior; (2) adaptive behavior converging into morphostasis; (3) adaptive behavior running through different phases of metamorphosis and morphostasis; (4) nonadaptive second-order behavior (generative).

fluctuating stories of trials and errors, of successes and failures, that evoke our own experiences of learning as fallible and imperfect entities. For instance, as I watch an agent such as a cart-pole system for an extended period of time, as it balance an inverted pendulum, discovering its environment, reaching plateaus of apprenticeship, I might start perceiving the desires of the agent, what it “wants”. I start to look ahead with apprehension, projecting both myself and the agent into the future.

I want to end with a few disclaimers. First, the “orders” of behaviors that I described should not be read as a hierarchy. Second-order behaviors are not in any way better or worse than behaviors of lower order: they are just different, they come with their own strengths and weaknesses, and can each be used efficiently (or badly) in artmaking.

Second, these categories are porous. For example, some mapping functions, such as moving averages or delays, have a short “memory” and can thus be said to have a “state”; some self-organizing adaptive systems have very limited structures which do not allow them to adjust significantly in

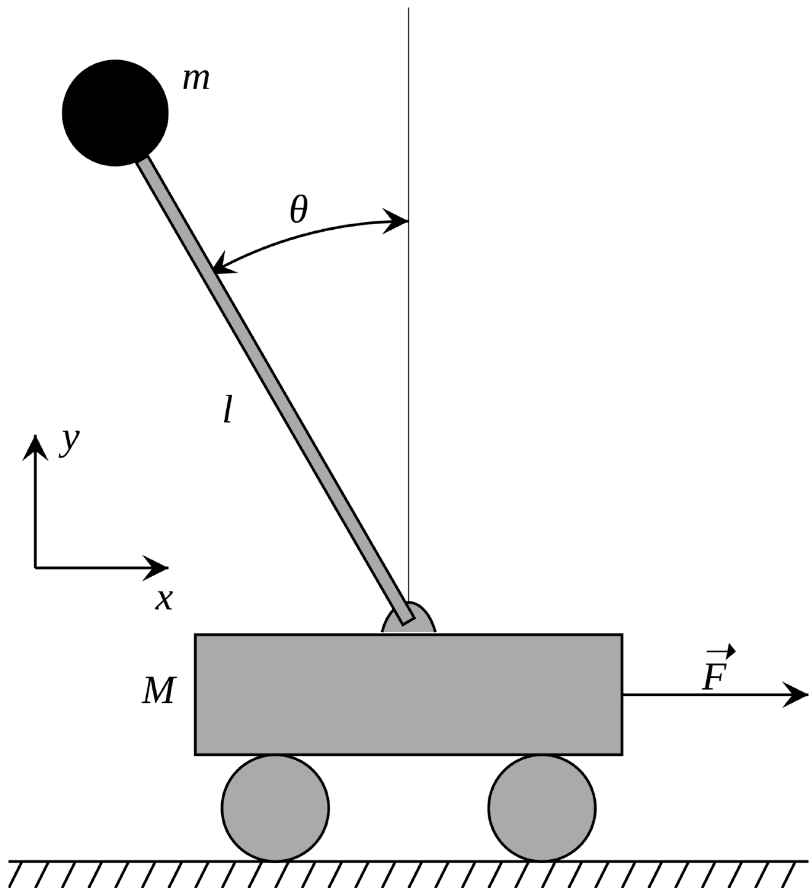


Figure 37: The inverted pendulum (or "cart-pole") is a classic problem for RL. By moving the cart from left to right, the agent needs to keep the pole in balance vertically, which is difficult because it is an unstable equilibrium. Source: Wikipedia.

the face of changing environments.

Finally, these categories can (and should, when appropriate) be mixed together. Most agent-based adaptive installations actually bring together a mixture of different systems, staging different kinds of zero-, first- and second-order behaviors, intertwining phases of morphological stasis, genesis and transformation intervening at different rates.[13] The use of lower-order behaviors gives the artist more direct control over the outcomes, which is often crucial for the success of a work.[14] For example, chapter 4 describes the swarming robotics installation *Vessels*, which integrates a formal-robotics structure known as a Behavior Tree and adaptive systems (Genetic Algorithms) in a distributed, evolutive choreography of autonomous agents.

This categorization is not meant as a systematic classification scheme, but rather as a frame of reference, a flexible analysis tool for artists and theorists. It gives an angle, a way to think and discuss about agent-based systems in art practice, that I hope can contribute to the language of new media as practitioners attempt to imagine new experiences and communicate their views with their peers.

5.13 Conclusion

This chapter focused on the notion of behavior in agent-based systems in general, and more specifically in works of art. I examined how properties such as embodiment, emergence, autonomy, adaptation and learning play out aesthetically in behavioral patterns, and proposed a set of theoretical tools with which to understand them.

I began by introducing fundamental notions related to agent-based systems in general, looking at a strand of research that took its origin in Cybernetics and GOFAI but ran in parallel to the history of Machine Learning which was explored in section 3.1. This historical analysis brought forward two significant considerations for a study of behavior aesthetics:

[13]Indeed, while recent research in the field of robotics suggests that the use of Machine Learning in robots is key to the advancement of the field, it seems to work better when used in combination with rule-based systems, at least at this point in time. In most studies, learning is used as a way to refine hand-coded processes or perform specific pattern recognition tasks. (Quinlan 2006; Chalup, Murch, and Quinlan 2007)

[14]This is, in essence, Downie's argument: he critiques both mapping (0th order behaviors) and emergence (2nd order) in favor of authorship in the design of programmed agents (of the 1st order). (Downie 2005)

1. A behavior consists in a pattern of events generated by an agent, as it is perceived by an observer who experiences it through its own sensory surfaces.

2. Behaviors do not exist as purely informational constructs, but are rather produced through the interactions of situated, embodied entities with their environmental surroundings.

New media art seems to have been much more influenced by this stream of research, which includes Artificial Life and Nouvelle AI, than by Machine Learning, at least until now. The key concepts of autonomy, embodiment, and emergence that run through it have been taken over by a number of artists and media theorist interested in the creative potential of such technologies as ALife and Nouvelle AI, which both seem to answer Roy Ascott's call for a new "behaviorist art". Yet, it seems to be disconnected in large part from the question of adaptation and learning in computational agents.

In an attempt to fill that gap, I examined Cariani's taxonomy of systems which brings together emergence, adaptation, and evolution in a unified framework. I connect it with Simon Penny's aesthetics of behavior and Soler-Adillon's take on emergent and self-organizing systems, looking at how adaptation and emergence play out in the generation of behaviors. In particular, I examined the temporal aspect of adaptive behaviors through the evolution of their morphology. I suggested to classify behaviors under three classes: (1) nonbehaviors (also called mappings, or zero-order behaviors); (2) behaviors (of the first order); and (3) metabehaviors (or second-order behaviors). The latest category involves processes whereby the agent's behavior itself undergoes morphological changes in time. These transformations include (1) morphostatis; (2) morphogenesis; and (3) metamorphosis.

I explained how adaptive processes pertain to the category of second-order behaviors and display morphogenetic patterns that lead to morphostasis once the agent has converged to optimum. Metamorphosis can be more or less abrupt, happening across different time spans, which influence the experience of these systems. Moreover, most successful artistic installations use a melange of approaches, possibly bringing the agent through different stages of evolution and stability, in an effort to generate a specific experience for the viewer.

In the next chapter, we approach these questions through a conclusionary analysis of the artistic work *N-Polytope: Behaviors in Light and Sound After Iannis Xenakis* (2012) on which I collaborated. This work rests largely on real-time generative behavioral patterns, some of which employ Machine Learning methods. As *N-Polytope* employs both adaptive and nonadapive approaches, it will act as a kind of test scenario for the proposed aesthetic theory of behaviors. I describe the process by which they were elaborated, looking at questions of practice. I analyze their outcomes, testing the application of the key concepts developed so far, examining the differences between the adaptive and nonadaptive methods used from the point of view of both the artists and the audience.

Chapter 6

N-Polytope

> A complex sound may be imagined as a multi-colored firework in which each point of light appears and instantaneously disappears against a black sky. But in this firework there would be such a quantity of points of light organized in such a way that their rapid and teeming succession would create forms and spirals, slowly unfolding, or conversely, brief explosions setting the whole sky aflame. A line of light would be created by a sufficiently large multitude of points appearing and disappearing instantaneously.
>
> – Iannis Xenakis, *Formalized Music*

In this chapter, I examine the installation/performance work *N-Polytope: Behaviors in Light and Sound After Iannis Xenakis*. A 2012 work created by Chris Salter in collaboration with myself, Marije Baalman, Adam Basanta, Elio Bidinost and architect Thomas Spier, *N-Polytope* brings the audience into an immersive, spectacular experience: a reinterpretation of famous composer Iannis Xenakis's series of large-scale media installations known as the *Polytopes*. The work received a special mention at the VIDA Art and Artificial Life International Awards and an honourable mention for the Prix Ars Electronica in 2013.

Focusing on the algorithmic dimension of the piece, I examine three different adaptive procedures that were used as ways to generate spatial patterns that unfold in time in a particular way. I compare these approaches to a nonadaptive algorithm using the morphological behavioral aesthetic framework introduced in the previous chapter, looking at the fundamental temporal qualities that differentiate adaptive from nonadaptive patterns in the work. I contextualize my analysis in terms of practice by investigating the work of Iannis Xenakis, his use of stochastic processes, and how he was influenced by Cybernetics. In particular, I oppose the notions of *command* and *autonomy* in Xenakis' work (Solomos 2006), contrasting it with works by early British cyberneticians and with our own approach in designing *N-Polytope*. I explain how Xenakis, while calling upon stochastic processes which were revolutionary for his time, guided by a notion of randomness, was indeed obsessed with keeping a strong, overseeing control over his pieces, a trait which directly influenced his vision of the relationship between art and science. Finally, I show how my approach to adaptive systems in artmaking can be interpreted as a hybrid between Iannis Xenakis' and John Cage's perspectives on *indeterminacy*.

In this chapter, I start by explaining the inception of the work, focusing on the relationship between the work of Xenakis and Cybernetics and explaining how this inspired our own work in *N-Polytope*. I give a short, high-level technical overview of the piece and then proceed to examine the core algorithmic realizations — contrasting adaptive and nonadaptive processes, looking at how they were used and what effects they had — using the aesthetic framework developed in the previous chapters. Finally, I conclude by addressing questions of control and time in adaptive agent-based installations, discussing important issues about audiences' expectations in traditional presentation contexts of new media art.

6.1 Xenakis, Cybernetics, and the Polytopes

Xenakis used stochastic methods for the first time in his 1956 piece Pithoprakta. He created the score based on a model of gas particles' speed and densities known as the Maxwell-Boltzmann distribution (Xenakis 1992). Here, gas particles were replaced by pizzicato glissandi sound grains whose steepness corresponded to the velocity sampled from the random distribution, resulting in a cloud of swarming sounds filling up the air.

Stochastic laws provided a way for Xenakis to generate compositions out of indeterminate systems, extirpated from directive human control. However, they still respond to macroscopic statistical laws and are thus still determinate at the global level, in the sense that they are statistically predictable.

The main interest of Xenakis in using these probabilistic approaches thus seems to be linked to the notion of entropy, which represents the degree of disorder of a system. He is interested in the potential for these distributions to generate "highly improbable events" that can result in sudden, "explosive" deviations from the average. While entropy is a concept borrowed from physics, it is especially interesting to notice it was one of the building blocks of early cybernetician Claude Shannon's information theory.

In his 1948 paper *A Mathematical Theory of Communication*, Shannon formulates the basic problem of communication as the transmission of a message from one point to another, going through a noisy channel (Shannon 1948). He posits that information is a quantity that corresponds to the minimum number of bits needed to encode a message. Showing how highly predictable messages can be encoded using less bits of information than unpredictable ones, Shannon points out that the amount of information needed to encode a message is directly related to its degree of unpredictability, which is mathematically equivalent to the entropy of the message.

Shannon's work had a tremendous impact on the way digital communication systems would be developed in the 20th century, as well as in a number of disciplines ranging from neuroscience to quantum physics. The simplicity of the model, its strong mathematical ground and its immediate applicability in the development of communication technologies of the time all contributed to its phenomenal success. Shannon's impact was so strong that his model gave rise to a whole new research field that ostentatiously labelled itself "information theory". However, one should notice that early cyberneticians were largely aware that Shannon's immensely reductive definition could only account for a facet of what information could be.

In fact, in his definition of Cybernetics, Wiener was referring not only to the Shannionian concept of information, which stays in the purely syntactic realm, but also to the semantic aspects of messages and their potential role in the control and regulation of machines, humans and society.

Indeed, Wiener coined the term "cybernetics" in reference to an 1867 article on feedback regulation mechanisms in boat governors from James Clerk Maxwell, the physicist who originated the theory of gasses (Maxwell 1867). "We have decided to call the entire field of control and communication theory, whether in the machine or the animal, by the same Cybernetics, which we form from the Greek *kubernetes*, or steersman. In choosing this term we wish to recognize that the first significant paper on feedback mechanisms is an article on governors, which was published by Maxwell in 1868, and that governor is derived from a Latin corruption of *kubernetes*." (Wiener 1961, 11–12)[1]

Wiener was interested in Shannon's theory because it provided a probabilistic framework for understanding how messages circulate between a system such as the brain and its environment. Yet, he was even more interested in how such systems would relate to their environment in a purposeful — one could say, meaningful — way by engaging in a self-regulating exchange with it. In this context, information acts as an active code that takes part in a control loop. Such systems, in their initial state, are highly entropic. Information flowing through negative feedback introduces a "degree of order (control)" since "information reduces uncertainty and contributes to order." (Smith 1974, 3)

An exception exists in some of his late works with algorithmic music composition. Xenakis used a self-organizing system known as cellular automata to produce the scores of his 1980s works *Ata* and *Horos* (Solomos 2006).[2] Cellular automata are "discrete dynamical systems with simple construction but complex self-organizing behaviour" (Wolfram 1984) which were first defined in the 1950s by von Neumann as a way to study artificial self-reproduction (von Neumann 1966) and extensively studied in the 1980s by people such as Christopher Langton (Langton 1984, 1986) and Stephen Wolfram (Wolfram 2002).[3]

According to Xenakis specialist Makis Solomos, the composer's idea of automata is related to Wiener's concept of *autonomy*, a vision that emphasizes the self-organization of the systems, as

[1] Interestingly, Maxwell's theory of gasses was an important inspiration for Xenakis. For example, in his masterpiece Pithoprakta (1955—1956), each instrument was acting like a molecule obeying the Maxwell-Boltzmann distribution, which models the speed of gas at a given temperature. (Xenakis 1981b, 29)

[2] Cellular automata were introduced earlier in the work of Artificial Life pioneer Christopher Langton. See section 5.2 .

[3] For the sake of clarify, notice that standard cellular automata are deterministic and thus cannot be considered stochastic processes. However their properties can make temporal structures emerge that makes some of them good candidates to be used as pseudo-random generators (Wolfram 1986).

opposed to the *command* perspective of von Neumann which is rather linked to a militaristic, "black box" model of machinic control. However, Xenakis does not fully endorse the idea of autonomy and seems to be interested mainly in using cellular automata as a tool that helps him shape his scores but over which he still maintains high levels of authorial control. Solomos explains: "Xenakis' manual interventions are very important; sometimes they destroy the nature of cellular automata. And, of course, they are far away from the idea of something that works alone, of an automaton, from which an autonomous meaning emerges." (Solomos 2006, 16)

We can only conjecture as to the reasons behind the relative absence of such autonomous systems in Xenakis' work up until the 1980s. The most logical explanation comes down to a question of means and intention. As an artist, he was interested in massive spectacular works that demonstrated complex, electrifying tensions between chaos and order such as the *Polytopes* and the *Diatope*. The kind of self-regulated, autonomous algorithms existing in the 1980s were just too simple and "toyish" to fulfill that vision. Furthermore, during the 1960s and 1970s eras, where Xenakis was the most prolific, connectionism and other forms of self-regulated systems were marginalized in favor of symbolic AI. Finally, the technology available to Xenakis was not advanced enough to allow real-time processing of large-scale input-output systems. Picture how most of his works were using cutting-edge resources, often running on university servers overnight, using FORTRAN programs encoded on perforated cards. The use of these systems by Xenakis as offline compositional tools thus comes at no surprise: there was really no other way.

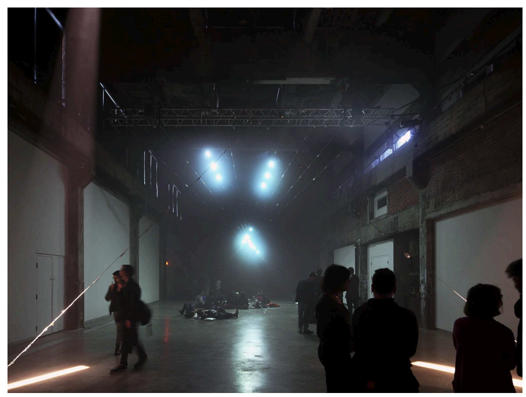
As artists, this was our point of departure in creating *N-Polytopes*. We reflected on the kind of work Xenakis would have created if he had access to the technology we have today, which permits massive real-time computation involving large-scale arrays of sensors and actuators. We posited that his interest in stochastic processes actually originates from an interest in the complexity of nature in both space and time. We thus supposed that, following the changes that happened in the scientific world from the 1980s onwards, he would have been more interested in the behavioral and complexity dimensions of Cybernetics and not only in its informational and probabilistic aspects.

In particular, however, we were interested in questions related to how emergent systems are felt by the audience, as well as an exploration of the continuum between order and disorder. The

(a) LABoral (Gijón, Spain) (2012).

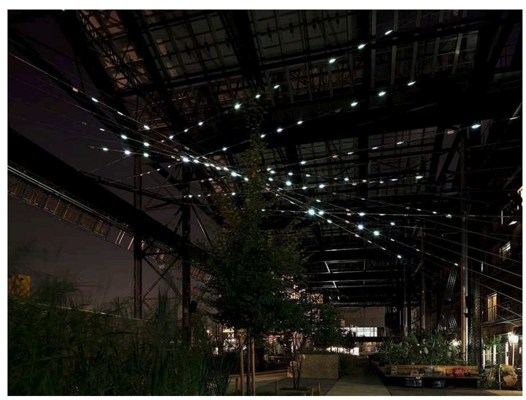
(b) LABoral (Gijón, Spain) (2012).

(c) Darling Foundry (Montréal, Canada) (2014).



(d) Darling Foundry (Montréal, Canada) (2014).

(e) Vitra Design Museum (Weil am Rhein, Germany) (2014).



(f) Nuit Blanche (Paris, France) (2015).

Figure 38: *N_Polytope* (2012) as shown in different venues. All images by Thomas Spier.

work of the late neuroscientist Walter Freeman on the brain as a dynamic system was an important source of inspiration (Freeman 1995, 2000). Freeman claims that brains function most of the time in a noisy manner except when cognitive acts take place: then neurons self-organize and random patterns settle into some kind of order. We were interested in the idea of phenomena that appear impermanently at the edge of order: the minute you think you have recognized them, they are gone.

In order to approach this, we worked with different kinds of real-time stochastic processes known as Markov processes. Named after Russian mathematician Andrey Andreyevich Markov, these random processes are characterized by their memory-lessness, meaning that the probability of future states depends only on the current state of the system (and not on the full history of preceding events). Markov processes are an important part of Xenakis' work and formalized music theory (Xenakis 1992). The most important extension to Xenakis' practice in *N-Polytope* lies our use of Markov decision process through reinforcement learning. A Markov decision process is a kind of Markov process that takes into account actions and rewards as part of the decision of a stochastic, memoryless agent, whereas reinforcement learning is an approach to address the decision problem when transition probabilities and rewards are unknown (Sutton and Barto 1998).

One of the main consequences of our approach is its impact on practice. Working with such autonomous devices transforms our relationship with matter, as it puts it out of our direct control. This suggests a reply to Xenakis's claim about art and science. As I already argued, Xenakis was not a scientist: he was an artist who used science and mathematics as a means to an aesthetic end. Xenakis recognized the growing importance of opening the dialogue between art and science. In a 1981 seminar at IRCAM, he proposed the establishment of a new relationship between art and science in which art would "pose problems that mathematics should solve" with the creation of new theories. He claimed that the "artist-designer" should be trained in various scientific fields ranging from mathematics to genetics, humanities and history, so that he acquires a kind of "universality" based on forms, architecture and morphology (Xenakis 1981a).

Xenakis' claim that artists need to come up with problems for science reveals, in my opinion, an asymetric relationship that does not match my own experience and vision as an artist trained in science. I suggest instead a relationship where science and art are willfully engaged into an embodied dialogue that goes both ways. In my experience, while it is true that scientific methods are used to respond to artistic problems, it is equally true that artistic questions are more than often inspired by scientific techniques. It is not a unidirectional process, but really a process of coupling, a performative endeavor that does not exist as a rationally organized communication system with clearly defined compartments, but rather as a blurry, embodied dynamics involving negotiations between matter, techniques and agents.

In the next section, I provide some details about the technical dimension of the piece, focusing mostly on the broad components and how they interrelate. This will provide some background for understanding the different processes that were implemented as part of the work.

6.2 Technical overview

Before we proceed into the heart of the chapter, let us first briefly sketch out the technical aspects of the work. *N-Polytope* consists of twelve (12) steel cables forming a customizable topological surface that is adapted to each venue. This architectural choice was not arbitrary but based on Xenakis' interest in projective geometry and ruled surfaces. On each cable are attached four (4) modules (or nodes) equipped with a Minibee board, an AVR-based microcontroller coupled with an XBee wifi technology developed by LabXmodal.[4] These modules form an ad-hoc network and can thus be controlled wirelessly from a central computer, either individually, as a group, or as sub-groups. Each module controls three (3) LEDs (light) and one speaker (sound) and observe data about their surrounding through a photocell (light) and a microphone (sound). A total of fourty-eight (48) nodes can thus be activated simultaneously, controlling up to 144 LEDs using pulse-width modulation.

Following Xenakis's own *Polytope de Cluny* (1972), a set of Minibee-controlled lasers are distributed across the space in specific spots together with both fixed and rotative mirrors, enabling the creation of geometric projections of concentrated colored light. A smoke machine can be activated to plunge the room in a hazy atmosphere that diffuses the light, contributing to the sense of

[4] See: http://sensestage.eu.

immersion and enhancing the lasers definition and intensity.

The piece can run in two different modes. In the *installation* mode, it employs more ambient and subtle sounds and light effects, without much evolution. Using a reduced set of algorithms, it aims to bring the spectator into a contemplative and reflective mood. Once every hour, the piece switches to its *performance* mode which runs for about 15 minutes. That version is more orchestrated, with a pre-rendered soundtrack and pre-defined cues that trigger other light and sound events, activating the hazing machine and the lasers, launching the different artificial agent-based algorithms across the cables and space, bringing the audience in more of a Wagnerian *gesamptkunstwerk*.

The algorithms, which were designed by myself, are used to activate light and sound spatially at a very coarse level, while a computational layer allows for their refinement in real-time before the final rendering. I examine some of these algorithms in more detail in the next section.

6.3 Adaptive Processes in N-Polytope

We now move to the core section of this chapter. Here, I review the computational composition of *N-Polytope*, focusing on four different agent-based algorithms which were used in the construction of the behavioral patterns that appear in the piece. The first procedure is nonadaptive and will be used as a point of comparison with the two other procedures. The second uses a simple adaptive feedback procedure applied to generate emerging synchrony between a large group of agents, whereas the third and last use reinforcement learning as a way to spawn dynamic patterns of light.

6.3.1 Drunk

The first algorithm that we implemented was named *Drunk*. It is a global-level agent that outputs, as its actions, intensities for all of the 144 LEDs using a hierarchical, one-dimensional random walk procedure. Random walks have been studied for more than a century. They represent a simple process where a variable changes its value by taking small, random steps, resulting in a staggering motion (hence the name “drunk”). Random walks are a kind of Markov process, a class of memoryless stochastic algorithms used by Xenakis in many of his works where the distribution of future states depends solely upon the current state and possibly a definite number of steps in the past, but not on the entire history of past states the algorithm has gone through (in the case of a random walk, the next state is only dependent on the current one).

The most simple way to apply this procedure to *N-Polytope* is to make a random walk in the 144-dimensional space of light intensities. However, there are other ways to do it as well, by grouping LEDs in different fashions and doing a random walk for each group. This is what was done in the case of *Drunk*. We created three different kinds of groups: 48 node groups (each controlling their 3 LED), twelve (12) line/cable group (each controlling 12 LEDs) and one (1) global-level group (controlling all 144 LEDs). These groups are mixed using four (4) parameters (one for the individual LEDs, one for the nodes, one for the cables and one for the global-level), allowing us to have a fine-grained control over the light environment generated by the procedure. For example, by setting the parameter controlling the global-level to 100% and the others to 0%, all the LEDs will have the exact same intensity which will stagger according to the random walk. If we set both the cable-level and the LED-level parameters to 50% and the others at 0%, we will see each line of LED staggering approximately at the same pace, with small variation for each LED. By setting all parameters to 25%, we get a mix of the four levels of control. We can think of this procedure as a hierarchical set of nested agents, with a global-level agent setting the general motion while the sub-level agents refine this motion down to the level of individual LEDs.

Thus, this procedure can be said to transform through time, bridging together past, present and future in the most simple of ways. It is not merely a mapping of something else, but rather a state-based behavior that displays some form of structured randomness. However, that structure does not evolve through time: observing the behavior after a minute, an hour or a month gives a similar impression. While the image is dynamic, constantly changing, the behavior itself is as still as a photograph.

Hence, applying the morphogenetic framework proposed in the previous chapter, we can classify Drunk as a very simple first-degree behavior. It is therefore morphostatic, since the patterns it generates do not change in time.

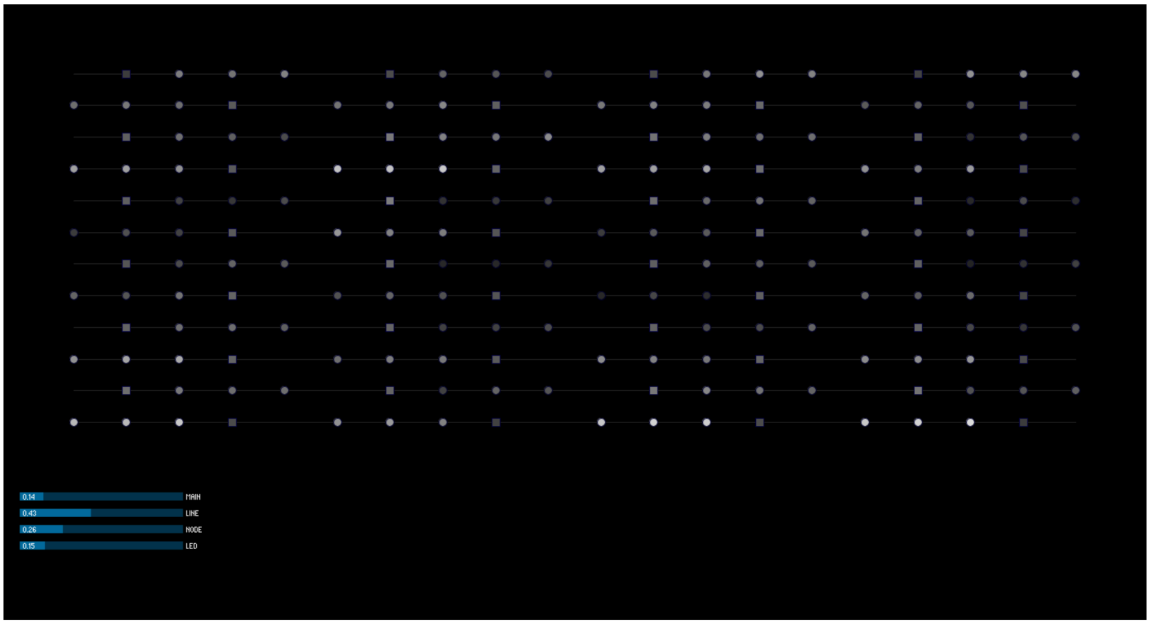


Figure 39: Screenshot of a GUI for displaying and controlling the *Drunk* algorithm. Controller nodes are represented as gray square boxes whereas LEDs are shown as round dots. The four control weights are manipulated using sliders in the bottom-left corner.

6.3.2 Fireflies

The second algorithm, called *Fireflies*, is rather different in nature. While the *Drunk* procedure does not have any inputs and is thus not influenced by the environment, agents in *Fireflies* are sending and receiving simple signals in the form of either light or sound. The algorithm is based on a procedure inspired from the behavior of fireflies in South-East Asia (Tyrrell, Auer, and Bettstetter 2006), originally designed to synchronize ad-hoc wireless networks.

In our implementation, each node of the *N-Polytope* installation is a "firefly" agent that "flashes" the LEDs it controls at a specific rhythm. When the procedure is launched, the agents are desynchronized: they flash at approximately the same frequency, however they are out of phase, resulting in sparkles of light that are emitted all across the room in random patterns, seemingly unrelated. However, each time an agent perceives — using its photocell — a flash of light coming from one of its neighbors, it adjusts its phase by a little bit. In other words, when it sees another flash, it adjusts itself to try to better match the flashing of its neighbors in the future.

As time passes, we begin to perceive an interaction between the agents. Some agents seem to be starting a chain reaction, where their flashing triggers their immediate neighbors, which in turn activate their neighbors, generating cascades of light bursting in massive ripples. Two, sometimes three such networks form, linked through space but separated in time. As they tend to refine their synchrony, suddenly, there is a long, dark silence. The room is filled with emptiness and time is frozen until, in a spectacular tempest of cold, white light, all the agents start flashing at the same time.[5]

This *Fireflies* algorithm works with either light or sound. At the beginning of the installation's 12–16 minute compositional cycle, we employ the same *Fireflies* technique on the audio component of the piece, which is synthesized directly from the 48 microcontroller nodes with similar results. In the case of audio, the synchronization is less perceptually discernable because the agents' sounds are mixed within the global background soundscape. However, the propagation effect of sounds responding to one another creates a strong immersive impression : the effect is uncannily similar to that of walking in a field of cicadas, or hearing frogs singing in a pond.

The procedure evolves from a chaotic, distributed, microscale behavior to a disciplined, singular, monumental one. Agents here have inputs (photocell/microphone) and outputs (LEDs/speakers) which are related through a feedback loop that pushes them to act the same way, in other words, to become a singular entity. Thus, one could say that as the system becomes increasingly ordered, it also becomes more monolithic. What is interesting from an aesthetic point of view is clearly not the purpose of the agents (which is to attain this perfect synchrony) but rather the process of going from discord to unison, which happens through the emergence of ephemeral temporal patterns. As we did in *Vessels*, we are hereby revealing the stochastic, adaptive procedure in its real-time unfolding as a way to create an aesthetic effect.

These agents can be said to be adaptive in the cybernetic sense, since they are engaged in a feedback loop where their own actions have an impact on the environment, which in turn influences their own future actions. Hence, the patterns they generate at the global level are different from

[5] I first used this algorithm in the media installation *Trace (S)* (2008) produced in collaboration with Jonathan Villeneuve. In that piece, a set of twenty-eight (28) cube-shaped devices hooked to the gallery wall synchronize using LEDs and photocells while pulsing air using computer fans.

those of *Drunk*, because there is an evolution in the light and sound that unfolds through time, embedded in a narrative of optimization. The behavior of the system as a whole is of second order. What the observer gets to experience is not only a certain real-time, recognizable pattern (like *Drunk*) but rather, a pattern which is a metamorphosis of patterns, transforming dynamically as it moves from chaos to order.

A relationship can be established between these very simple units and neurons. Indeed, the algorithm driving *Fireflies* is a form of "integrate-and-fire" oscillating neuron (Campbell, Wang, and Jayaprakash 1999) whose origin can be traced back to early 20th century French neuroscientist Louis Lapicque (Abbott 1999). This category of time-based, asynchronous[6] neural models is widely different from Perceptrons and MLPs, which are synchronous by definition and are targeted at pattern recognition rather than biological simulation and temporal integration. In the field of neural computation, these models are often called Spiking Neural Networks (SNN) (Maass and Bishop 2001), a connectionist model of biological neurons that adds a conception of time. In (Thivierge and Cisek 2010), the authors describe how such models can be trained to recognize and generate synchrony. Here the evaluation function is the time difference between an observed flash and the phase of the agent, and the update rule consists in slighly changing the phase so as to match it more closely with the perceived phase of the flashing neighbor.

6.3.3 Boosters

The third algorithm, called *Boosters*, works in a similar fashion to *Fireflies*, although it utilizes RL techniques: there is one agent for each light and sound node, each of which can emit flashes of light as well as perceive the brightness of its neighbors. Agents in *Boosters* accumulate energy in a virtual "battery" while they are at rest, collecting the light emitted by their neighbors. At each step, the agents can choose to either stay at rest or emit a burst of light. If they choose the latter, the sum of the energy they have accumulated is spent to produce the burst with the intensity of the light emitted being proportional to the spent energy.

[6] By this, I mean that the neurons fire independently at their own rythm, however as we have shown it is possible for them to learn how to fire synchronously.

Boosters agents get rewarded for producing a flash, however, they receive an even larger reward for producing a more intense one. Their best strategy is thus to wait until their battery is full before taking a flash action. Since there is a "blind" relaxation period after emitting a burst during which the light perceived from the environment does not add to the energy, the agents' best strategy as a group is also to intersperse their flashing. From a visual point of view, the perceptual impression that results is one of a mass of individual lights pulsating over a range of different intensities only to occasionally burst and blank out for a moment in order to again return to their struggle.

The *Boosters*, however, seemed to be rather unable to learn even this apparently simple procedure. I suggest this could be attributed to the fact that the space of inquiry (in other words, the number of possibilities to consider for each agent) is too large for them to learn in real-time because they aren't exposed to enough data in the time during which they run.[7]

If that is true, being able to witness a learning process happening in real time using generic learning approaches such as RL would require that the problem to be solved by the agents stays relatively simple (small search space). That would, in turn, completely defeat the purpose of using ML techniques in the first place. The curse of dimensionality therefore creates a nasty situation for artists wanting to exploit such methods by building of aesthetic experiences through the staging of real-time adaptive behaviors.

6.3.4 Chasers

The last procedure, called *Chasers*, simulates agents moving across the installation's cable structures using a reinforcement learning algorithm combined with an artificial neural net. Here, instead of working with continuous properties like light intensity and sound amplitude, we are rather using a discrete representation of the agents' position in space. Each cable represents a one-dimensional "world" with twelve (12) discrete locations/cells. The world "wraps-around" at the end, meaning that the first cell is considered to be adjacent to the last one. At any specific moment in time, an agent occupies one and only one of the twelve cells and can choose to either stay in place or move

[7]Looking backwards, we could perhaps have managed to train them using an offline technique such as Batch RL (Lange, Gabel, and Riedmiller 2011).

to one of the adjacent cells. The only information (observation) the agent receives is the distance (in number of cells) between itself and the next agent, in both direction. The agents' positions are represented by lighting the corresponding LED on the cable (stacked agents result in a brighter light).

The reward function is the sum of three different components:

- The reward on touch (r_t) rewards the agent (or punishes it) for being on the same spot as another agent.
- The reward on move (r_m) rewards the agent for moving in a given direction (and punishes it for going the opposite way).
- The reward on stay (r_s) rewards the agent for staying put (and punishes it for moving).

These parameters can be used independently (by keeping the other ones to zero) or they can be combined to foster different behaviors in agents, as demonstrated in the following table:

Table 5: Example reward functions for the *Chasers* procedure, with corresponding expected results.

Touch (r_t)	Move r_m	Stay (r_s)	Resulting behavior
1	0	0	Try to catch other agents at all cost.
-1	0	1	Try to evade from other agents and otherwise stay still.
0	1	0	Move left-to-right no matter what.
2	-1	0	Try to catch other agents first and foremost, but with a preference for moving right-to-left.
-1	0	-1	Always move but avoid collisions.

Furthermore, we can achieve more variation by combining agents with different reward functions on the same cable, thus generating different kinds of movements such as predator-prey "chases" and other adaptive dances. In the installation, we start by adding a few agents and allowing them to stabilize, which happens rather quickly. We then manipulate the tension between chaos and disorder — an important aesthetic dimension of the work — in two different ways.

First, we took advantage of a feature of the RL optimization procedure which was introduced in section 3.2.2: the exploration-vs-exploitation criterion. A parameter $\epsilon \in [0, 1]$ controls the probability that the agent will, at any given step, "explore" its environment by taking a completely random move, as opposed to "exploit" its current knowledge by taking a "greedy" action (i.e., moving in the direction it thinks will yield the highest reward). Choosing a low value for ϵ typically yields more structured, "smart" moves (especially after the agents have been given enough time to learn) while a high value will generate chaotic behaviors. By playing with ϵ we can influence the behavioral shapes visible to the human audience, moving them instantly between order and disorder.

The second strategy is used when agents have "stabilized" into an ordered, "smart" behavior, entering a phase of morphostasis. By increasing the number of agents in the space, we make the problem more difficult and confuse the agents that are already there. It also raises the density of light sources activated along the wires, until the structure becomes saturated both spatially and temporally, shifting the patterns towards chaos. In this sense, the shift in the behavior of the agents gradually results in a growing sense of disorder, achieved by sudden discontinuities in the rhythm of their movement up and down the lines and thus, making it increasingly difficult for observers to recognize their patterns.

In the various contexts in which *N-Polytope* was presented, the audience seemed to be dragged into the piece, often describing it as compelling and hypnotic. Many viewers stayed for extended periods of time, lying down under the structure, engulfed into the cosmic spectacle offered by the piece. *Chasers* and *Fireflies* seemed to be particularly effective as methods for inspiring this general awe in the viewer, for what I believe to be two different reasons.

What made *Fireflies* compelling was that one could experience the actual learning process, the slow adaptation of agents to one another where they auto-organized in real-time. This was particularly effective as it happened at the beginning of the sequence, allowing a slow ramping up towards higher energy movements. As *Chasers* happened in a later, climatic segment of the show, it was tuned to learn faster, and was also augmented with additional light effects that made the behavior less "pure". Due to this it was harder to see the learning process happening because the agents quickly stabilized into an optimal pattern, and the definition of their movements was blurred by the added spectacular effects. Nonetheless, the rapid movements of the agents underlying their actual representation follow behavioral patterns that, although hard to discern, are somehow evocative of recognizable shapes.

6.3.5 Discussion

This report described a number of different approaches that make use of adaptive and Machine Learning algorithms as part of agent-based artworks. In particular, it shows one instance of using reinforcement learning with success[8] as part of a media installation. At the broad level, the strategy is similar than that employed in *Vessels* for the co-adaptation of behavioral patterns using genetic algorithms, in the sense that we can play on two different levels.

First, by hooking into the adaptive loop, we can unfold it before the eyes of the audience, revealing the learning process itself as a behavior. This is particularly true of *Fireflies*, which opens the show by allowing groups of agents to self-organize temporally, moving from chaos to order. Here, the time span is much shorter than for *Vessels*, the adaptation happening over a period of 2–3 minutes (compared to 10–15 minutes in *Vessels* for a noticeable change to happen). The result is impressive, as one gets to experience the real-time, incremental adjustment of light and sound agents surrounding them, building up a dramatic transformation from individuality to unison.

Reinforcement Learning in *Chasers* goes even faster, as the agents converge after about a minute. So the strategy differs from *Fireflies*: agents learn an optimal behavior, and it is this stabilized, morphostatic performance that is revealed before the eyes of the audience. In *N-Polytope*, this seemed to work out, although it is not absolutely clear why given the triviality of the learned behavior. The software layer built on top of the behavior generation system, which augments these temporal patterns with diverse fine-tunings and effects at both the local and the global levels, makes it difficult to analyze what aspect of the general experience can be attributed to the learning algorithm itself.

So it seems that we are confronted again with the same paradox. On the one hand, the kind of problems that are interesting aesthetically are often very high dimensional and complex. Hence, these are also the problems for which Machine Learning is useful, as compared to more traditional

[8] In fact, *n-Polytope* is, to my knowledge, one of the rare documented works of art employing RL, along with some of my most recent works *Fifth Absence* (2011) and *Plasmosis* (2013).

approaches. However, the context of artistic works that run in real-time is improper for training these systems under the time constraints offered by traditional new media presentation contexts. In the kind of scientific applications which they were designed for, Machine Learning algorithms are usually trained offline, learning off disk-based databases, running as fast as the computers on which they run can crunch the numbers. Naturally, this kind of computational power is not yet fast enough or small enough to drive a live artistic event such as *N-Polytope* which runs over a time span of about 15 minutes in performance mode.

Though it leaves a number of questions open, this research has at least the merit of highlighting important distinctions between adaptive and nonadaptive algorithms. Stochastic processes such as the multi-stage random walk implemented in *Drunk*, which is rather emblematic of the kind of algorithms Xenakis was interested in, exist in a domain that is essentially distinct from adaptive algorithms such as *Fireflies* and *Chasers*. While they could all be said to pertain to the category of behavior-based aesthetic, they certainly differ in their temporal unfolding. Whereas *Drunk* is a temporal process that creates a strong impression in the eyes of an external observer, the structure that supports it does not adapt or evolve over time. Despite its strong temporal dimension when considered over a short period, it is still flat and static when taken from the perspective of a longer span of time.

Fireflies and *Chasers* pertain to another category of behaviors. Because they can adapt their structure over time, and that this structure in turn determines their behavior, their relationship with time is of a very different nature. The GA employed in *Vessels* for the evolution of the behaviors is similar, however, we could even make a further distinction between algorithms such as the one in *Vessels*, which changes behavior over time but without a definite goal, from *Chasers* and *Fireflies*, which have precise goals.

Applying the morphogenetic framework proposed in the previous chapter, we can classify *Drunk* as a very simple first-order behavior. It is therefore morphostatic, since the patterns it generates thus do not evolve in time. In comparison, the three other algorithms are behaviors of second order, or metabehaviors, because their dynamic shapes evolve in time. *Chasers* and *Fireflies* are both characterized by an adaptive optimization narrative, their behavior being iteratively constructed in time (morphogenesis) until the system reaches an stable equilibrium (morphostasis). However, in the case of *Chasers*, this process happens much faster, almost instantly, so we do not really get to see the adaptive process. Here the changes and the way they are impacted by learning are more blurry, as there are many variables that come into play (the variation of ϵ and the adding of agents which impact the environment in which they evolve). *Fireflies* thus offers, in my view, a much more “pure” example of the aesthetic effects of an adaptive metamorphic process involving embodied agents. Furthermore, the behavioral properties of *Fireflies* allow it to learn over a short time span of a few minutes, giving the audience an opportunity to experience the synchronization of adaptive units in real-time.

Audiences respond to the piece with a feeling of enthrallment. It is not rare, during presentations, to see people staying for extended periods of time, ranging from tens of minutes to hours, which is rarely seen in new media artworks. Many people have described that the patterns unfolding before their eyes seemed elusive, effervescent, always on the verge of being grasped, then dissolving. Trying to make sense of what steered he system, what were the entities behind it, audience members would make often turn up to analogies of living systems to make sense of the patterns observed in the work: swarms of insects, flocks of birds, shoals of fishes, moving and scintillating in a way that felt lifelike, yet never fully graspable. It may be possible that the force that kept these audiences in thrall was that of being witness to alien modes of being and behaving that lie beyond human understanding.

6.4 Control and Time

Temporal morphological characteristics of processes such as the ones that were explored in *N-Polytope* inspire important questions about the use of adaptive agents in new media art installations. What aesthetic effects do these morphological movements activate? How does working with partially uncontrollable and indeterminate agents affect practice?

As a way to approach these questions, I will use as a complementary example Stephen Kelly’s piece *Open Ended Ensemble* which was briefly introduced in section 3.2.1. Kelly is one of my collaborators on *Vessels*, and has a background in both art and science, as he is currently completing a Ph. D. in computer science at Dalhousie University (Halifax, Canada) in the field of Genetic Programming. The latest version of his piece involves two agents that adapt concurrently, a strategy similar to the one I used for my 2013 underwater installation *Plasmosis*. The first agent consists in a set of four pairs of neon lights that can choose to switch off one (only one) pair of fluorescents. The second agent controls inaccurately the movements of a magnet moving near the fluorescents, trying to find the spot with the lowest magnetic "noise". The sound of the magnetic fields is directly transmitted through a guitar amp of the artist's making.

I interviewed him in June 2016 about his work which was displayed at Hamilton Artist Inc. in Hamilton (Canada). Asked about how the audience experiences the piece, and how Machine Learning plays into it, he says:

> Working with such technologies involves a loss of control on the part of the artist, a strategy which is very common in media art these days. And by doing that you are re-introducing elements of uncertainty in a piece, you are giving it the potential to surprize you. Something is controlled, but not by the audience nor the artist, so in a way it is more "democratic", it lies outside of the control of anyone. And that in turn speaks to how people experience the work: you cannot understand it right away, even if you read about the piece, and even as its designer. You need to get to know it. So in that sense it is a very durational experience. It is common in many media art works but even more so with adaptive systems.

As a point of comparison, Kelly mentioned the research of Adrian Thompson, a scientist at the University of Sussex who is considered to be a pioneer in the field of evolvable hardware. In 1996, he used a Genetic Algorithm to evolve a circuit known as a Field-Programmable Gate Array (FGPA) to discriminate between a 1kHz and a 10kHz tone (Thompson 1996). After the circuit evolved into an optimal discriminator, in order to simplify it, he tried to prune out the parts of the circuits that were not contributing to the output. Contrary to Thompson's assumptions, some parts of the circuits that were completely disconnected from any path that could influence the output, were actually crucial to the discrimination process, probably through some forms of local magnetic interactions. In other words, the adaptive agent that controlled the evolutionary process had learned a solution to the problem that made use of the intrinsic, embodied, physical properties of the circuit and that no human could possibly have come up with.

Similarly, human observers of adaptive or evolutionary works such as *Open Ended Ensembles*, *Vessels*, or *N-Polytope* cannot understand their behaviors rationally, because the underlying processes that govern them follow non-logical rules. Works that are based on mappings and first-order behaviors can be rationally explained and understood: for example, this photocell triggers that sound effect, that microphone activates that video sequence, that gestures causes the agent to start running in circle for a minute, etc. But in order to experience second-behaviors in all their richness, one needs to "get to know them" phenomenologically, through her own sensorimotor body. One needs to adapt to it, to change herself, to become attuned to it until the behavior reveals physically itself in all of its unfathomable nature.

I posit that this is a direct consequence of two important features of adaptive systems that have been highlighted earlier. First, the way by which their morphology evolve in time may contribute to their mystifying nature. As was pointed out before, adaptive systems go through periods of morphogenesis, metamorphosis and morphostasis. During their transitive phases, their behavior lies in a state of flux, making its shape difficult to grasp by external observers.

Second, as they become better at performing the task they are trained for, Machine Learning models grow into complex and intricate architectures that are more than often unintelligible to humans. This is particularly true of neural networks, where large numbers of independent units (neurons) work together to solve a problem, yet, it is difficult if not impossible for a human observer to find out which neuron is responsible for what, because decisions are usually diffused across the network. Adrian Thompson's genetically evolved FGPA is another good example of how Machine Learning agents find their own way through problem-solving, often moving beyond human logic.

Incidentally, this structural complexity is precisely what allows these systems to be efficient, evolving intricate behaviors that often make them perform better than humans. It is thus not surprising that the morphological patterns emerging from these architectures (once they stabilized) remain perplexing to human observers.

More importantly, one should not forget that the human observers experiencing these systems

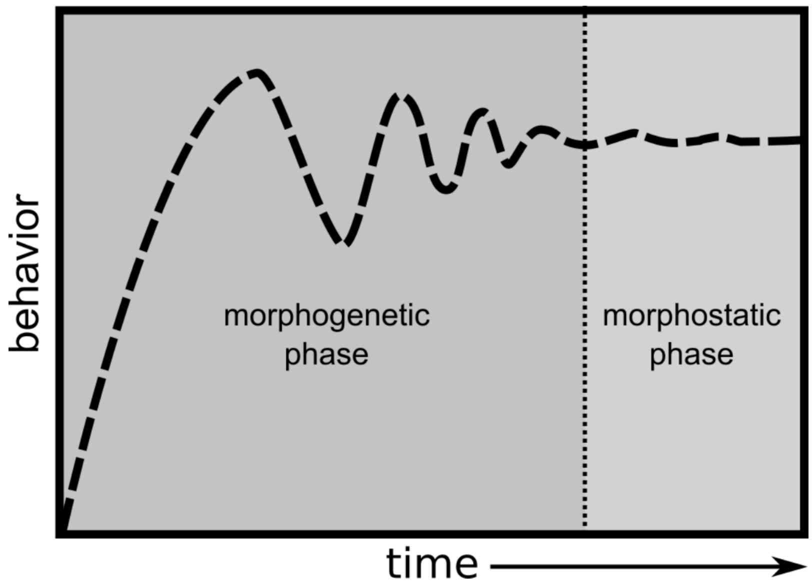


Figure 40: Two properties of adaptive behaviors might contribute to their unfathomable nature: (1) the unstability of their behavior during morphogenesis/metamorphosis; (2) the complexity of their behavior when it stabilizes into morphostasis, due to the complexity of their trained model.

and trying to make sense of them are, themselves, adaptive agents. The process by which human observers encounter behaviors is one by which they tentatively and iteratively adjust their expectations so as to better predict the movements of the agents, becoming familiar with them, getting to know them. In the case of morphostatic behaviors, for example, it is possible, given enough time, to become intimately attuned to them. If the behavior is of first-order, governed by a set of logical rules, it is theoretically possible that she could then be able describe these rules, or at least get a sense to know them rationally. However, if these behavioral patterns are governed by an intricate self-organizing structure such as a neural network, one will still not be able to describe the experience in terms of explicit rules (because there are no such rules!), yet, might be able to intuitively "feel" the intentions of the agent producing these patterns — in other words, to "understand" them phenomenologically.

An important consequence for media arts is that the contexts appropriate for these systems, and in particular the expectations of the public, need to be transformed in order for these works to be brought to their full potential. This is particularly complex as the audience of media art has been trained to interact with mapping and formal systems, which can be grasped quickly. As always, it is difficult to convince a large public to make an effort to understand an artwork, and most people will leave a work that they cannot grasp within a frame of a few minutes.

Since adaptive systems require a lot of data to learn how to react to complex situations, they usually require a lot of time, enough that they can have access to many different experiences. In this way, they are very similar to complex living systems. Consider for example the time it takes for pets and even humans to learn about things that seem relatively simple, such as walking. Strategies can be deployed to prevent the audience from leaving too early. For example, in *N-Polytope*, a number of nonadaptive effects are used to immerse the audience into the work, such as a smoke machine and an underlying music score. The audience is informed that the piece runs in an installation mode followed by a 15 minutes performance mode, which makes them want to stay long enough to "see everything". A set of carpets and comfy bean bags are dispersed under the structure, providing a very simple yet effective way to increase the time spent with the work.

But this also suggests that the new media art scene is perhaps not currently adapted for these works. The new aesthetic experiences these pieces offer necessitate a change, both in the way new media institutions present and mediate them to the public, and in the expectations of the audiences. It is possible that audiences have become accustomed to a certain type of new media artwork that works well within a festival-oriented art network, and that might have been fostered in parts by the popularity of mapping and interactive art in the last three decades. Long-form, evolution-centric works such as *Vessels* and *Open Ended Ensemble* places value on duration and contemplation, making them more akin to performance art practices where similar concerns of embodiment and performativity come into play.

Furthermore, aside from these changes, we need to create new contexts that would allow these for these new forms of experience to unfold naturally, such as take-at-home agents that would evolve inside one's living quarters over weeks, or public art monuments that would adapt to their environment over the scope of many years, allowing the public to develop a relationship with these systems so that they can "get to know them" and experience their adaptiveness.

6.5 Indeterminacy

It is commonly known that visionary composer Iannis Xenakis was inspired by the aspect of Cybernetics that concerns the probabilistic nature of information, which would later be known as Claude Shannon's information theory. It is less known that his interest in Cybernetics extended to Wiener's notion of control in autonomous systems.

Quite interestingly, while Xenakis was keen to employ probabilistic methods in his work, demonstrating his fascination with chaos and order, he was also extremely demanding, and controlled each aspect of his work with extreme precision in a top-down manner that is related to von Neummann's *command* perspective on Cybernetics.[9] In realizing *N-Polytope*, we have been leaning more towards the Wienerian notion of *autonomy*, allowing our agents to move freely, working with a much more bottom-up approach.

This was both an effect of vision and means. As an artist working with technology, my own practice has always been defined by the *mise en scène* of artificial computational agents, and autonomy is a defining feature of my behavior-based aesthetics. This, of course, largely influenced our decisions in the kind of algorithms that were implemented as part of the work. It was a response to the question: "What would have Xenakis done if he had access to the technology that we have?" The technology Xenakis was working with could not generate things in real-time: everything needed to be pre-rendered, which suggested making changes to the score, because there was time to react. That limitation has been largely alleviated nowadays, allowing for large-scale real-time media processing like was used in *N-Polytope*.

This is not to say that *N-Polytope* is a purely autonomous work. The behaviors that were described in this chapter were integrated inside a traditional score, with predefined cues that triggered the various events that gave form to the piece. This is particularly true in the performance mode

[9] In a personal conversation during the presentation of *N-Polytope* at the Darling Foundry in Montreal (2014), Robert Dupuy, who worked as an assistant for Xenakis on the Polytopes, seemed to suggest this. Recalling the times where he worked for the artist, he explained that, as a programmer, he would write a program and run it, generating a new score — a process which would take 24 hours. Then Xenakis would look at it and ask for some specific and precise parts to be changed, such as one or two notes that he would not find satisfactory, requiring Dupuy to perform yet another round-the-clock iteration. Dupuy's story seems to support the claim that while Xenakis was using stochastic systems as part of his creative process, he also took a strong authorial position, not allowing for much autonomy or control in regards to the machines.

of the work, which runs for a defined period of time. Nonetheless, while the performance itself is predefined at a higher level, the artificial actors that create the performance are never doing the exact same thing, which contributes to the natural feel of the piece.

The investigation of *N-Polytope* in the current chapter shows the advantages as well as the limitations of the tools developed in this dissertation. In particular, when working with emergent systems, a tension always exists between the artificial systems, the audience, and the artists. While I agree with Downie and Xenakis in reaffirming the importance of authorship when working with self-organizing agents, I dissent with them in regards to the way to enact that authorship. Both Downie and Xenakis emphasize the role of the artist as an arbiter, with an unequivocal right to make choices among the various contents generated by computational processes. In the case of Xenakis, this seems to come down to an almost divine right, which is well rendered by the composer's final words in his thesis defense. Replying to Bernard Teyssèdres who asked him about the importance of the ability for the composer to "select" preferred sonic versions generated by stochastic programs of his design, as opposed to leaving them out of his control, he stakes:

> But it is my right, my privilege. It's my task to prefer one thing over another. (Xenakis and Messiaen 1994, 98)

This tension between autonomy and command in Xenakis' work echoes his conception of *indeterminacy* and the impact of Western philosophy and modern science on his work. Musicologist Kostas Paparrigopoulos has contrasted Xenakis' approach to chance and disorder with that of experimental composer John Cage. Both Cage and Xenakis made use of indeterminate processes in their work, but in a different fashion. While Xenakis aimed to use science as a way to control chance and shape it to its will, Cage saw in the unpredictability of nature an opportunity to break free, leaving sounds out of his direct control. (Paparrigopoulos 2011)

Cage's reasoning is based on the observation that sound exists beyond human intention. Describing his visit into an anechoic chamber at Harvard University, he explains he could still hear the sound of his circulatory system and a very high-pitched sound which he claimed came from his nervous system.[10] "Until I die there will be sounds. One need not fear about the future of music."

[10]Cage most certainly did not hear his nervous system in that chamber. What he might have been hearing is

Cage continues:

> But this fearlessness only follows if, at the parting of the ways, where it is realized that sounds occur whether intended or not, one turns in the direction of those he does not intend. This turning is psychological and seems at first to be a giving up of everything that belongs to humanity—for a musician, the giving up of music. This psychological turning leads to the world of nature, where, gradually or suddenly, one sees that humanity and nature, not separate, are in this world together; that nothing was lost when everything was given away. In fact, everything is gained. (Cage 1961, 8)

Given this conundrum, the composer thus has two options. If he refuses to let go of his effort to control sound, he can do as Xenakis, seeking complex techniques as ways to better approximate nature. Otherwise, he can choose to "give up the desire to control sound, clear his mind of music, and set about discovering means to let sounds be themselves rather than vehicles for man-made theories or expressions of human sentiments" (9).

Free will being incompatible with determinism, and freedom being the prime substance of originality and creativity, both Cage and Xenakis wanted to use indeterminacy as a path to escape their own individuality and channel larger powers in the universe, to "go beyond themselves". While Cage, inspired by eastern philosophies, sought to attain this freedom through the abandonment of control, giving their full autonomy to sounds, Xenakis, in his attempt to deconstruct the determinism-indeterminism opposition, always stayed on the side of western philosophy and science. (Paparrigopoulos 2011, 3)

These divergences resonate with many of the dilemmas that have followed us throughout this research: determinism versus indeterminism, command versus autonomy, computationalism versus enactivism. On the one hand, the instruments and methods of science allow us to create representations of natural processes that can grant us control over them. On the other hand, we are confronted by the fact that nature is out of control, that we can never fully know it, and that our best shot at dealing with it is to boldly accept it as it is, or, as Cage says, to "let things be themselves".

My own perspective in regards to these quandries faced while working with computational

tinnitus, which often results in a high frequency sound. (McElhearn 2016) behaviors stands somewhere between a Cagean and a Xenakisian approach to indeterminacy. My interest in Machine Learning as a way to generate self-organizing dynamics in agent-based systems resonates with Xenakis' usage of stochastic science as a way to bend randomness to his own will. However, echoing Cage, I want to leave the agents "be themselves" and "live their artificial life", so to speak. Still, there is more of a blurring between human and inanimate in my work than in Cage's work, who wants to limit as much as possible human intervention to allow chance to take its course. I see my role as an artist less as a director around which everything else revolves, and more as a collaborator negotiating with other human and nonhuman agencies in the production of artworks.[11] As an alternative to rule-based AI, Machine Learning — and, in particular, Reinforcement Learning — is situated in the middle ground between these two visions, as it allows for a certain degree of control over outcomes by giving the artist the right to design goals, while leaving the learning agent the responsibility to find the way to reach them. There is a huge difference between the use of coin flips (in Cage) vs a system that is based on experience. There's a difference between a set of rules vs a system that learns by experience, vs using a system based on randomness.

I recall here Andrew Pickering's nonmodern ontology, which conceives the world as "built from performative dances of agency". Pickering remarks, following Latour, that modernism is built on an "asymmetric dualism" that considers humans are the only beings able of true agency, thus positioning humanity above other animals and things.

> But then comes the twist. The great discovery of science studies was that in practice the sciences themselves fail to exemplify this ontology. It turns out that in their own laboratories the scientists are far from calling all the shots. They do not dominate their materials through knowledge; instead they engage in rather symmetrical open-ended and performative *dances of agency*, trying this and that in their struggles with machines and instruments, *finding out* what the world will do in this circumstance or that, and responding to what emerges in a process that I call *mangling*. [...]
>
> So studies of scientific practice conjure up a new ontological vision, a quasi-biological one, of the world as itself as a lively place, itself a reservoir of agency, that can always surprise us in its performance, and that we always have to *get along with* and accommodate ourselves to, rather than seeing through and controlling. We are always, so to speak, in

[11] Indeed, collaboration has always been an integral and important dimension of my art practice, a particularity which directly stems from my background in science, where collective work is the norm rather than the exception.

the thick of things. (Pickering 2009, 198)

This idea directly resonates with my own artistic approach, and provides a strong argument for addressing the problematics that this dissertation brings forward in trying to understand the worldview that adaptive systems come to generate. Art is not meant to provide answers or to be a vessels of communication, but rather a space of encounter, a theater where there different agencies co-adapt: the work, the artist, but also the audience. So perhaps this is what adaptive systems suggest when used in the arts: both in terms of experience and practice, they are the quintessence of life's unfathomable nature: open-ended, muddy, indeterminate processes, which we can never fully know, let alone control.

6.6 Art and Science

This discussion suggests a review of Xenakis' perspective over the relationship between art and science. Xenakis claimed that while both art and science could construct objective forms of knowledge by drawing out ideas through empirical observation and rational thought (*inference*) and putting these ideas to test using experiments (*experimentation*), only artists had the power to attain subjective truths through the process of *revelation*. Xenakis proclaimed the advent of an artist-engineer who would need to be "simultaneously rational (inferential), technical (experimental) and talented (revelatory)" (Xenakis and Messiaen 1994, 5–6) in order to implement a merging of art and science.

Xenakis sees this artist-engineer as an interdisciplinary researcher and practioner, trained in a wide range of scientific and artistic fields. He argues in favor of a new relationship between science and art where the artist-originator would invent artistic problems that she would then try to solve using mathematics and science (Xenakis 1981a). He claims that the numerous failed attempts at making music using computers at the time were either due to musicians' relative ignorance of basic notions of mathematics, physics and acoustics or to scientists' general lack of points of reference when it comes to aesthetic creation. To him, art and science form an alloy, an heterogenous substance whose properties are different than those of its constituents (Xenakis and Messiaen 1994).

Xenakis's vision is very noble: that through artists creating problems for science to solve, art and science are forming "alloys" with each other possessing new emerging properties (similar to that of bronze, which has other properties than copper and iron). This being said, Xenakis' vision fails to account for the sociopolitical context in which these field interoperate, in particular within the millenial context of socioeconomic power relationships between art, science, and capitalism.

For example, the corporatization of academia has utterly transformed scientific research since the 1970s. The contemporary context marked by an extreme pressure to succeed in their disciplinary field makes it difficult for scientists to get involved in projects involving interdisciplinary research with artists because the return on investment remains unclear. There are almost no economic incentives for scientists to shift into the arts. In counterpart, artists moving into scientific environments are still subject to the financial and power unbalance described above, and are thus at risk of finding themselves either used as contributors that can "think outside of the box" for a cheap wage, or be forced to abandon their art career in favor of a more sustainable job as scientists or engineers. Xenakis' idea that artists should lead humanity's quest for truth by creating questions for science to solve thus seems hardly imaginable in the current situation.[12]

The performative ontology brought forward by Andrew Pickering in *The Cybernetic Brain*, explored in section 5.11, suggests an alternative to Xenakis' alloys that I find more fruitful (Pickering 2010). In his book, Pickering turns towards early British Cybernetics as an alternative approach to modern science, which engaged directly with matter in often strange ways, such as through the creation of bizarre apparatus such as Ashby's *homeostat*, Walter's tortoises, and Pask's *Musicolour*.

> At the most obvious level, synthetic brains—machines like the tortoise and the homeostat—threaten the modern boundary between mind and matter, creating a breach in which engineering, say, can spill over into psychology, and vice versa. Cybernetics thus stages for us a nonmodern ontology in which people and things are not so different after all. (18)

[12]Perhaps Xenakis, an immensely successful architect and composer sitting at the top of the world, had trouble seeing the power relationships induced by the forces of capitalism the way most artists do. Perhaps he was able to pull these forces his way: but an exception is no rule, and more than often, scientists nowadays either do not have time for artists, or when they do, it is more than often in situations where the power relationship is unbalanced in their favor. This asymmetry in economic conditions is one of the social factors that prevent Xenakis' alloys to happen in the contemporary world. In a way, one can see how Xenakis' alloys are a tribute to an idealized modernist vision of art and science that fails to take into account the socio-historical reality in which they are embedded, and the diverse actors that participate to their construction.

Following that perspective, Xenakis' claim, that the emotional and affective dimensions of truth are only accessible by art, seems dubious. The role of intuition in science has been well documented, and it is well known that the emotions, passions, and motivations play an important role in science. Scientifics are human beings as much as artists, and in their work they touch a unique beauty that is often hard to transmit outside their circles, whether it is in the astuteness of mathematical proofs, the complexity of chemical reactions, or the elegance of biological interactions.

This is why I find that Xenakis' approach is not only inaccurate (because I think there is a lot of subjectivity in science, perhaps as much as in art), but it is also barren, as it precludes alternative forms of relationships to take place beyond the idealization of an artist-engineer who puts science to the service of art. If we adopt, instead, a perfomative vision of both art and science practices, where practitioners and researchers engage in a "dance of agencies" with matter, then it becomes possible to imagine different kinds of relationships. For example, it allows for artistic strategies where an artist discovers a scientific technique that holds an aesthetic potential and brings it out of science — such as many of Marcel Duchamp's *readymade* works like *Fountain* (1917) and *Bicycle Wheel* (1951). This is similar to what I am doing in this dissertation, and in much of my work: I am often inspired by a technique that holds a potential to reveal some subjective truth and generate novel experiences, and I extirpate it from its scientific context, deconstructing it to bring it in the field of arts.

As an alternative to Xenakis' artist-engineer figure that is more compatible with Pickering's view, consider Simon Penny's historical analysis of "artist-inventors" in his 2008 text "Bridging Two Cultures – towards a history of the Artist-Inventor". There, Penny tries to position interrelations between scientists and artists in technological practice, focusing on technological practitioners engaged in the design of "machine-artworks" capable of generating embodied interactive experiences with the audience. On the opposite side of media users who often see technology as an end itself, these artist-inventors rather develop the technologies they need to achieve their objectives. In contrast with Xenakis' figure of the artist-engineer (who is an artist trained in engineering), Penny's artist-inventors category comprise people who are traditionally considered more like scientists than artists, such as Cyberneticians Ross Ashby and Grey Walter, and engineers Nikola

Tesla and Alexander Graham Bell. As thus, it is a much more inclusive concept which provides an alternative view against the modernist separation of art and science. (Penny 2008)[13]

6.7 Conclusion

In this chapter, I used the large-scale performance/installation work *N-Polytope* as a testbed for the aesthetics of adaptive behaviors I have articulated through this dissertation. Of keen interest is the comparative usage of both adaptive and nonadaptive systems in *N-Polytope*. The morphological framework was useful in describing behavioral evolution in the different kinds of processes. The various algorithms appearing in the work exploited the tension between morphogenesis, metamorphosis, and morphostasis.

The experience of the work as described by members of the audience echoes in many ways how people experienced *Vessels* (c.f., section 4.5), seeing patterns that evoke alien agencies, pursuing behaviors that are hard to pin and can hardly be described. This, in a way, is not surprising, when one considers how observers are, themselves, adaptive systems, who are trying to make sense of often complex behaviors that might or might not be stabilized yet. These self-organized systems need a particular context to be correctly apprehended: they cannot be consumed like your usual reactive, interactive art piece. One needs to spend time with these agents to get to know them — in other words, to adapt to them.

Hence, adaptive systems and their use in agent-based artistic installation point to a view about art and science that questions the notion of indeterminacy. In this context, I see the artist less as one who attempts to give an idealized shape to matter by controlling randomness through science,

[13]This is certainly not to say that science and art are equals. As Penny claims, whereas "many of the innovations in science and technology, arose from a passionate commitment to specific causes or ideas" and while the "drive to invent and the drive to create are, at root, almost indistinguishable", it remains that scientists are "taught to discount motivations which exceed the positivist quest for knowledge, while artists have no such constraint" (Penny 2008, 143) In an earlier text, Penny compared how artists and scientists approach computer-based artworks. For example, he contrasted the very scientific approach in the Oz Project — a research program led by computer scientist Joseph Bates in the mid–1990s, that aimed to develop technologies that would allow artist to design complex dramatic interactive fictions — which is very generic and complex, but engendered, in his opinion, rather limited results, to the much more low tech approach of artist Luc Courchesne which yields a more evocative and achieved artwork in his piece Family Portraits (1993). Penny pretends that scientists, due to their training, tend in general to focus on problem solving and to produce "universal" solutions which favor didactic and literal works, whereas artists focus on the experiental aspect of the work in its relationship with the audience and the environment, thus focusing on the poetic and metaphorical aspects of the work (Penny 2000).

but rather as one who tries to negotiate her way through trial and error, and exploration and exploitation, in an interconnected network of both human and nonhuman agencies. In this regard, I feel that the artist is ideally participating in the morphological evolution of her agents, yet is keenly aware of her imperfect control over them, and in fact uses that awareness as a powerful transformative instrument.

Adaptive Computation and Machine Learning provide ways for artists to design new behavioral patterns, not by controlling each and every aspect of the outcomes, but rather by effecting the intentionality of agents who are then left out of any direct human control from either the authors or the audience. Authorship thus becomes much more diffuse, and experiences are brought into a domain of uncertainty that evokes the meeting of an alien form of life. Understanding these new experiences and the practices they are associated with demands a change of perspective from both the artists, the audience and the art institutions. Human-centered principles of control needs to be replaced by a posthumanist view that takes better account of how both human and nonhuman actors intervene in the artistic process through adaptive/performative movements of resistance and accommodation.

Chapter 7

Conclusion

> My intention is to let things be themselves.
>
> – John Cage

Through this study, I set out to explore the concepts of adaptive systems and Machine Learning in agent-based new media installations. The relevance of this pursuit is demonstrated by the uniqueness of its subject of inquiry, the originality of its methods, and the significance of the questions it seeks to address in the field of contemporary media art, as summarized by the following premises:

1. Adaptation plays a central role in contemporary conceptions of life, bridging the gap between self-organizing phenomena and living systems. Rooted in adaptive principles, the field of Machine Learning has been steadily growing in presence in the contemporary landscape of computer science since the early 2000. After half a century of research, Machine Learning finally seems to have succeeded in surpassing the popularity of rule-based approaches to AI.

2. Adaptation is a core principle that is necessary for understanding scientific and artistic practice. For example, Pickering's concept of a "dance of agencies" and Xenakis' notion of art-science "alloys" both articulate adaptive as well as learning mechanisms that resonate with work in Cybernetics, Artificial Life, and Machine Learning.

3. Because of its capacity to engage critically and creatively with matter and experience, art can uniquely contribute to the understanding of these concepts, generating alternative narratives and imaginary grounds for future learning agents.

4. There seems to be a discrepancy between, on the one hand, the growing importance of the field of Machine Learning in AI since the mid–1980s and, on the other hand, the relative scarcity of artworks that make use of adaptive systems, not to mention the lack of both practical and theoretical frameworks for understanding such works. The research described in this dissertation provides historical context as well as practical information for addressing this gap, for example by developing an aesthetics of adaptive behaviors as they appear in embodied agent-based art installations.

Following the principles of art-as-research, the study also highlights my experiences employing an iterative research-creation methodology inspired from Agile software design. Seeking the development of new knowledge through the aggregation of material practice and theoretical analysis, I opened up this research by examining the Absences series of experimental intervention in natural settings from 2008 to 2011. I described how installing agents in hostile and changing environments further highlights questions about adaptation, revealing the necessity of doing this research. As a way of better framing these questions, I followed the entangled historical paths surrounding the emergence and progress of Machine Learning and Adaptive Computation from the 1950s onward, examining in particular their influence on the development of media art. In parallel, I examined important concepts tied to these research strands, such as embodiment, enaction, coupling, autonomy, emergence, self-organization, authorship, and performativity. Through this theoretical research, augmented by reflexive accounts of practice derived from my experiences creating the artworks *Absences* (2008—2011), *Vessels* (2010—2015), and *N-Polytope* (2012), I address the research questions core to this research project:

1. What new forms of aesthetic experience do Machine Learning methods enable or make possible when utilized outside of their intended context and are instead carried over into artistic works?

2. What characterizes the practice of using adaptive computational methods in agent-based artworks?
3. What kind of worldview are these works fostering?

7.1 Experience

One of the biggest challenges of this dissertation has been to extract specific, focused observations for comprehending how human audiences experience the type of agent-based, adaptive computational artistic installations which are the focus of this study. This pursuit is complicated by the fact that art, and New Media art in particular, is extremely dense and diversified, and more often than not involves a multiplicity of technological and artistic strategies. This situation makes it difficult to isolate the specific effects associated to the use of Machine Learning techniques in their construction.

As a way to address the aesthetics of such systems, I considered their autonomous coupling with the real world through the notion of enaction, which provides a model for understanding novelty and meaning generation through an embodied agent's interactions with its environment. I examined how adaptation acts as the way by which self-organizing systems are able to generate novelty in the work of Peter Cariani, and is therefore an important mechanism in the emergence of creativity and life. Based on these ideas, I extend Simon Penny's concept of "behavior aesthetics" by providing a framework for understanding behaviors in terms of how their shapes evolve through time.

In this taxonomy, three kinds of behaviors were distinguished, namely: (1) mappings; (2) first-order behaviors; and (3) second-order behaviors, or metabehaviors. I claimed that adaptive and evolutionary systems are the only ones capable of producing this third category, because their behaviors themselves change over time according to an underlying metamorphic behavioral pattern. While this does not make these systems aesthetically "better" than others, it reveals how adaptation can generate different kinds of experiences deeply tied to their temporal unfolding.

To describe the ways by which these behaviors transform through time, I reference the notions of morphostasis, morphogenesis and metamorphosis. Behaviors of the first order are purely morphostatic: their shape does not change through time. Given enough time for their dynamic patterns to appear and repeat themselves, an external observer can recognize and familiarize themselves with them. Contrary to these behaviors of the first order, adaptive behaviors usually oscillate between periods of morphogenesis/metamorphosis and morphostasis as they adjust to their environment. If the environment and the goals they are trying to meet remain stable, they will eventually stabilize in a morphostatic behavior of the first order, until the conditions change again.

As such, the complexity of the self-organizing circuits that are trained in Machine Learning algorithms lay beyond human comprehension. This is especially true in neural networks such as MLPs, whose trained weights are difficult, if not impossible, to interpret by human beings (d'Avila Garcez, Broda, and Gabbay 2001). Adrian Thompson's mid–1990s experiment with a genetically trained programmable circuit described in section 6.4, is an excellent example of that problem. Once the FGPA was trained to distinguish between two audio frequencies, the scientist removed the parts of the circuits that were completely disconnected from the inputs and outputs, thinking that there was no possibility they could logically be involved in the computation. However, he realized that by doing so he destroyed the functionality of the system, which had learned to make use of the analogic variations induced on the signal through the disconnected yet physically present parts of the circuit.

The second-order behaviors induced by adaptive systems, combined with the inability of humans to understand their structural organization, are two important features of such systems. This may explain why audiences subjected to such works are often mesmerized by their uncanniness, as they are never able to settle on a familiar, recognizable morphology in the behaviors that unfold before their eyes. This is why these works require a period of adjustment from the audience, as only by spending time with the piece can they really begin "knowing" its behavior. This knowledge of the work is not one that of specific logical processes, but rather a knowledge of generalized behavior that one might become acclimated and perhaps become close to.

This process usually takes time and cannot be described using technical or precise languages, not even by the artist who designed it. Adaptive systems are thus evocative of alien agencies, of lifelike processes from another world: observing these behaviors is an underdetermined, subconscious, concretely unexplainable experience. What it implies aesthetically for both artist and audience is a certain loss of control in regards to plastic or logical footholds for aesthetic experience, and instead that the aesthetics might be found in behavioral patterns that evolve following rules that cannot be pointed to directly, but felt through one's own adaptive body.

7.2 Practice

As a first step in understanding the artistic implications of working with Machine Learning, I examined scientific definitions of Machine Learning. I looked at how the different components of a learning algorithms can be exploited by artists, namely: (1) models; (2) optimization procedure; (3) evaluation function; and (4) training data. I used this format to demonstrate how artists could work with ML methods to build self-organizing behaviors by focusing on each of these constituents as aspects of creation.

An important aspect of artmaking with ML that constantly resurfaces is the relationship between the complexity of the problem to solve, the capacity of the model (i.e., the number of free parameters that can be adjusted, such as weights in a neural network or DNA bits in genetic algorithms), the number of input and output dimensions and the quantity of data points available to the system. To make a long story short, complex problems usually require bigger and more powerful models, which in turn require more data. A direct consequence for artmaking is that, if one wants to generate an adaptive behavior happening in real-time through an embodied agent, she must either make the problem simple, or allow the agent to learn over a longer period of time. Even relatively simple problems can extend beyond the timeframe characteristic of New Media art presentation opportunities : problems can take days, weeks, or even years to solve.

These time constraints therefore complicate the trial-and-error process usually involved in the creation of such behaviors in New Media artwork,[1] which necessitates a certain loss of control. It is often impossible to have the system behave exactly the way it was imagined, and therefore the artist is often at the mercy of the very same system that she is designing.

Machine Learning offers a framework for working more efficiently with self-organizing systems, a

[1]Imagine having to wait for a year before verifying that an agent has learned something aesthetically satisfying, only to realize there was a bug in the system or that some other adjustment is required.

set of tools for "shaping" agents by assigning them goals and desires. Machine Learning thus holds the potential to facilitate the development of emergent systems when compared with "bottom-up" approaches such as those of ALife. Yet, in no way does it preclude the process of trial-and-error that one goes through when making art. In a way, the process is even more antagonistic than in more traditional artforms, because on the one hand, the artist often wants these systems to surprise even herself, to become alive, so to speak, independent of one's control; yet at the same time, as an artist, one wants to create certain effects, and the more the systems become alive and autonomous, the more they might drift away from one's will and generate patterns that might be difficult to understand.

Perhaps a solution to this conundrum is to accept that one cannot have both autonomy over these processes and command them at the same time. So in the end the creative process demands that one let down their guard and accommodate themselves to these processes, let the Machine Learning agents "be themselves", and then get to know them, and be changed by them. This is, of course, easier said than done, but my hope is that the accounts of practice and the various tools presented in this research help facilitate this process for artists wishing to work with adaptive and evolutive agents.

7.3 Worldview

This loss of control is key to the development of a posthumanist perspective over art and science in the 21st century. Adaptive behaviors suggest an alternative path between the mythical figure of the artistic genius who selects the "right" processes, and the capitalist utopia of artificial creativity. They ask for a new way to look at our relationship with other forms of agency.

Furthermore, the use of Machine Learning agents in artistic practice demands of artists and audiences alike to examine and perhaps revise their expectations about information, control, and time in regards to technologies themselves. The kind of artworks described in this thesis do not provide any information, they drift in and out of everybody's control, and they deploy over very diverse time spans. This suggests that the media demands new ways for relating to art, science, and life, beyond the consumerist, tractable, and efficient forms the Western world has become so accustomed to in technological interactions, towards undetermined, incontrollable, mutating, non-optimizable processes which have a true potential for change.

Peter Cariani's theory of biological semiotics, described above, highlights the profound adaptive nature of emergent behaviors in both living and computational systems. His characterization of the relationship between emergence, adaptation, and the production of novelty in such systems resonates directly with Andrew Pickering's perspective. Pickering describes how systems faced with a changing environment that becomes incompatible with their model of the world (resistance) have to adapt by redefining themselves (accommodation), thus suddenly diverging from their expected behavior.

Pickering turns to early British Cybernetics to find an alternative way to do science, in a way that both engages directly with matter and might fully embrace the indeterminate and muddy characters of research. From this, he developed the idea that scientific practice should be seen as an ongoing interaction between human and nonhuman agents, a "dance of agencies", considering scientific practice in a way which is decentered from the human subject. Pickering's view of scientific research might also suggest a means of addressing the demands placed on audiences and institutions by emergent technologies in 21st century New Media art. The complexities of adaptive systems can shed new insights on the relationalities between the artist, the audience, and the artworks. They demand from art theories that are able to account for the various human-machine interplays that happen in making and presenting art. In the context of agent-based systems, this suggests the artist should be considered not so much as the commander of her work, but rather as a collaborator with human and nonhuman agencies that literally develop the artistic object into its own intersubjective becoming.[2]

[2] In particular, this further undermines the romantic figure of the solo artist, which has been put into question by modern art but is still pervasive in the popular imagination.

7.4 Implications

The theoretical and practical implication of this resonate across multiple fields of research in the humanities. First and foremost, this research engages with contemporary discourses of computational and robotic systems in media art and STS by clearly articulating the power of adaptive and learning systems and behaviors within larger historical and conceptual frameworks. My research uniquely traces the impact of these concepts of adaptation through the history of both science and art alike since the post-war era. It shows the active part played by Machine Learning and Adaptive Computation in the history of Artificial Intelligence — a role that seems to be growing exponentially in the past decade — while simultaneously demonstrating its impacts in media art.

This historical study of adaptive agents has led me to develop a framework to help art theorists and practitioners come to a more cohesive understanding of behavioral aesthetics in agent-based artworks. The proposed taxonomy differentiates "mappings" from behaviors, and suggests a special category for the kind of "metabehaviors" produced by self-organizing systems in general (and adaptive systems in particular). I believe that this system might help one to better define behaviors of Machine Learning agents, taking into account the imagined perspective of an external observer that perceives recognizable patterns performed in the agent. I suggested the concepts of morphogenesis, metamorphosis and morphostasis, borrowed from biology, to describe the way such second-order behaviors change shape through time.

By bringing these conceptual tools into the contemporary discourse of media art, my goal is to provide artists and theorists with ways to imagine, express, discuss, and criticize works of art that makes use of computational (or computationally-inspired) processes. In particular, my research on adaptive and self-organizing systems suggests that emerging, more revolutionary forms of art are made possible through the use of agent-based systems that perform behaviors of higher order. It supports and extends the idea that artificial agencies described by STS scholars such as Pickering (early Cybernetics systems such as Ashby's homeostat and Walter's tortoises) offer the possibility of a new range of experiential effects through the deployment of behavioral patterns that evoke our own complexity, imperfections, and indeterminacy as embodied, living beings performing in the world.

Yet, it also suggests that the new media art scene might not be currently accustomed to works with highly adaptive systems, as new media art has developed as a field largely through exhibition contexts and festivals that acclimatize its audiences with mapping and clearly defined procedural operations. As such, artists working with adaptive systems might perhaps find more ready contexts for their practice by working within more traditional art networks that have long accepted subtle, long performance of embodied practice as defining aspects of their history: performance art, environmental art, and perhaps relational aesthetics. The other option is that institutions presenting digital artworks could adapt in order to allow for these works to be presented under the right conditions, to educate and develop the public into getting in the good state of mind to appreciate these works.

As adaptive systems have increasingly become instruments of total control in contemporary capitalism under the regime of “big data”, artists could perhaps subvert this idea of control and authorial command into creating technologies of unseated becoming and subjecthood. In particular, there is a way by which the lifelike behaviors of Machine Learning can become a space of emancipation from the art-as-object commodity drive of capitalism: if audiences are ready to accept artworks “as themselves”, the objects as a relational locus can provide a means of artistic interaction that resist logics of ownership and authorial control.

Further to the consideration of digital art, this study points to the need for a reconsideration of how artists are conceived of. If we admit that art has “never been modern”, that it is a blurry, muddy, performative dance that engages many agents, then the figure of the solo artist should hold less sway. If practitioners want to truly understand how art and science can work together in the creation of new aesthetic experiences, one has to also recognize how she is embedded in a network of agencies that include significant economic and political asymetries in a 21st century neoliberal context.

This in turn suggests a call for the deconstruction of the traditional notions of the solo authorial artist in favor of models that actively embrace collaboration and multiple authoring as aspects of contemporary creation, perhaps even assuming the model of the sciences in favor of formalized systems for co-authorship.

The relationship between arts and sciences cannot be conceived of any long as “pure” in a modernist sense — because they never were, as anthropologist of science would claim (Latour 1991) — but should rather be seen as human activities embedded in specific networks and ways of doing — engaged in material practice, negotiating with agents in an adaptive fashion, transforming themselves as much as they transform the agents they are working with. Media art in particular is hybridized, and for some artists, the practices of science and art are increasingly intertwined.

7.5 Limitations

This study of course is conducted in recognition of many limitations: I will hereby only discuss at length the limitations that have the strongest impact on the quality of current findings.

First, this research is based on the important premise that the choice of a particular algorithm to control the behavior of an agent has a direct impact on how it is experienced by the viewer. While the aesthetic framework that I introduced in this dissertation has the advantage of clarifying some of the conceptual and formal aspects of behavior as a medium for artmaking, the research did not go as in-depth as it could have in understanding the role of the observers in the process. Whether one uses mappings, first-order and/or second-order behaviors as part of an agent-based installation, the experience of the work can never be separated from the intricate relationships that are established between the viewers, the work, and the environment.

While much attention has been placed on understanding the intrinsic characteristics of learning algorithms and the evolution of their morphologies, the role of the observer in this process has been comparably asymmetric. It is still not clear how the usage of Machine Learning, as opposed to more simple techniques from Cybernetics or GOFAI for example, really make a difference when it comes to audiences—the development of large scale qualitative and quantitative research in this regard is outside the scope of this thesis.

The second major limitation of this research is the weakness of its methodology when it comes to evaluating the public response to adaptive works. Indeed, the first limitation is in large parts a consequence of this methodological flaw. My examination of these reactions could have been more objective and better documented, for example through a more thorough qualitative study.

More generally, the focus on algorithms over human interactions characterizes an important shortcoming, a lack of languages on behalf of researcher and audiences alike for how these computationally-generated behaviors are received beyond a vague sense of "lifelikeness" and "uncanniness". These relationships between artwork and observer are subjectively encountered, as it is only the observers themselves that have the power to assign agency to things around them. Social robotics expert Cynthia Breazeal considers that making artificial agents autonomous does not make them "sufficiently life-like", and argues that believability is an important aspect in the design of social robotic agents because it projects the "illusion of life" and gives the agent a personality (Breazeal 2002, 8). In this light, the revolutionary nature of this work only holds if it is capable of inspiring a consistent social impact. While this is true of most forms of experimental art, it is still premature to make statements of the effects of these works on audiences who are still developing.

Moving on, the artistic projects that I analyzed were complex assemblages of different technologies and physical objects, making it difficult to precisely analyze the impact of the algorithms used because they were intertwined with the aesthetics of their physical form as well as with other sensory media. A more systematic set of experiments, where one could isolate the algorithms and compare their effect in front of an audience, could have been beneficial in trying to tackle the aesthetic question.

The third main weakness of this study lies in the fact that I feel like I have only scratched the surface in regards to the establishment of a wider aesthetic worldview. It is still not clear how one can imagine a new relationship between art and science through a nonmodern ontology. Does the performative nature of practice adequate insist on a deconstruction of the frontier separating art and science? How do adaptive systems, and in particular connectionist and deep learning architectures, challenge both computationalism and enactivism in their conception of cognition, and how does this affect our view about representation and embodiment in art and science practices? How is indeterminacy connected to such a performative worldview? There are only a few of the many questions that still remain to be explored.

7.6 Future Work

The results of the study as well as its limitations suggest a number of future projects that could address the study's shortcomings, as well as extend its findings. Improved research approaches need to be enacted for addressing questions of experience and practice in audience members. Can we assess of human experience, and how? How to evaluate artificial agency and behavior morphologies from the audience's perspective?

The evaluation of experience presents important challenges. First, the concept of human experience itself is equivocal. When we use the term "human experience", I consider, along with thinkers like David Chalmers and Stevan Harnad, that we often conflate two distinct notions (Chalmers 1997; Harnad 2000). One meaning of "experience" refers to the set of functional, causal, and possibly observable events happening in the physical world during consciousness. For example, when we describe an experience using human language, or when we measure it using data such as biosignals, interviews, or questionnaires, or otherwise employ measures that seem to presuppose rules or general principles of human perception, we are referring to this tangible aspect of experience, which I will refer to as "experience-as-function".

The other meaning of experience concerns the phenomenal feelings associated with such events, only accessible by the conscious subject living through them ("experience-as-feeling"). One of the most fascinating characteristics of the world we live in is that, even though felt experience does not seem to be playing any causal role in it, the only thing we can truly know for certain is, as Descartes rightfully claims, that we are feeling. Yet, another fundamental principle of our universe known as the "other minds problem" is that we can only feel our own experience of the world: we do not have access to the felt experience of other bodies (Harnad 1991).

It is hence impossible to evaluate someone's felt experience, because measurement can only be done using a third person's perspective, which is forbidden by laws of our universe. Yet, unless one delves into solipsism, it is rational to assume that similar bodies, because they have close physical properties, when placed in the same context, subjected to the same conditions, and affected physically in similar ways, will likely feel something analogous.

This correlation existing between function and feeling is what allows us to use different methodological approaches to measure experience-as-function. A number of strategies have been deployed in the fields of user experience (UX) and human-computer interaction (HCI) that might be of interest. Each evaluation method, however, has its drawbacks: for example, if we ask the subject to describe her experience while she is living them, this will affect her experience, while if we ask her after, she might not be able to recollect fully how she felt.[3]

Thus, it is possible to evaluate human experience in works of art, assuming we understand that (1) what we measure belongs to experience-as-function, that (2) any measurement will be imperfect and incomplete, and (3) that any instrument we use for observing the subject will impact her experience and how she reports it. Experience is highly context-dependent, subjective, and dynamic (Law et al. 2009). The best approaches need to take into account the context of the work, in order to both limit the impact of the measurement method on the lived experience while getting the best results, and also to facilitate the establishment of correlations by subjecting audience members to a similar range of media events. This is very hard to do because works of art often have open-ended contexts, and the human subjects that see the work of art are subjected to many incontrollable affects that might be completely independent of the work itself.

Consider, for example, how evaluation of audience's experience could be implemented in the cases of Vessels and N-Polytope. These two works have two completely different contexts. Vessels is designed as a public artwork where a mixed audience of passers-by as well as a targeted audience can observe the ongoing spectacle of a community of artificial beings in public space. N-Polytope, in comparison, is designed for indoor presentation as an immersive experience that engulfs the whole experience of the audience members present. The piece also oscillates between an ambient mode and a performative mode (which contains a narrative progression).

UX metrics are traditionally categorized as either *external* (i.e., based on observations from an external examiner) or *internal* (i.e., generated by the user's own observations over her experience).[4]

[3]Francisco J. Varela has proposed a set of neurophenomenological methodologies that try to bridge the gap between first-person and third-person accounts of experience, but I do not find his argumentation satisfying (Varela and Shear 1999).

[4]Importantly, even "internal" methodologies cannot access felt experience directly: they only provide descriptions of these experiences, through the vantage point of the subject living them.

External evaluation methods would be unlikely to provide useful data in *N-Polytope* because very few observable events usually happen that could give hints about the user's experience. In the case of *Vessels*, it might be useful to use video and audio recordings to observe spectator placements and movements around the robots as well as conversations between them.

One category of internal methods that could work well for both *N-Polytope* as well as *Vessels* is the *thinking aloud* method, (Lewis 1982) where users are provided with a microphone and asked to describe their experience in real-time, in their own words. Since in both cases we not only want to collect information about the audience's global experience, but rather how the different behavioral patterns generated by the works comes to affect this experience in real-time, it would be useful to use some form of recording as a reference frame for both internal and external data gathering methods.

For example, in the case of *N-Polytope*, we could align the think-aloud session with the message cues sent by the software running the show. This way, we could know, for example, how the spectator reacts when the "fireflies" or the "chasers" patterns begin and end. In the case of Vessels, since there is no centralized system that sends the triggers, a video recording, as suggested earlier, would suffice to align the robotic behavioral happenings with the spectator's reaction.

Semi-structured interviews (Edwards and Holland 2013, 2—3) would also be an interesting method of measurement applicable to both cases, allowing the interviewer to investigate specific questions while leaving space to allow for unforeseen ideas to flow during the conversation. In particular, such techniques would be useful for *Vessels* because they could be applied more easily to the input of passersby — for whom it might be impractical to ask to be hooked up with a microphone for a think-aloud session. These interviews could be run on the spot with voluntary visitors, interrogating them on their experience while promoting dialogue with groups (e.g., families, couples, friends). For *N-Polytope*, it would be better that these interviews be made retrospectively, to allow visitors the opportunity to experience the sensory experience without interruption.

An interesting, complementary approach called experience workshops is described in Edmonds, Bilda, and Muller (2009). It involves bringing together a small group of experts from various fields of study in arts, social sciences, HCI, etc. After experiencing the work on their own, the group meets and each participant describes her experience, while the others take notes. They then work together on a set of questions about the work and generate a report.

New experiences need to be designed in order to better understand how people experience different orders and types of behaviors. *Vessels* and *N-Polytope* are aesthetically complex works, making it difficult to separate the influence of the behavioral patterns from the other media components such as sound and light. It would be useful to design specific experiences where (1) users would be confronted more specifically with the different types of behaviors, while keeping the media elements stable; and (2) another set experiences where the behaviors are kept stable but are realized through different kinds of media.

The appropriateness of different contexts for presenting agent-based works needs to be more thoroughly studied, perhaps by comparing the experience of a work when it is presented inside a traditional art venue, a new media festival, in public space, or in a semi-private space (such as a workplace that would share its audience's everyday life). More research also needs to be conducted adressing issues of duration, such as examining the audience's experience of adaptiebehaviors over different time spans. In relation to these questions, I would like to continue to explore the work that has been done in the field of social robotics.

An area I am interested in exploring further is the tension between a computationalist/representativist and performative/enactivist worldview. While I agree in most part with critiques of computationalism, I find the enactivist model to be limited. Like Harnad, I think that representations exist in the brain and play an important part in cognition, however, these representations are not, like the computationalists claim, pure symbols independent from a physical substrate, but are rather grounded in the body (through connectionist networks). I am interested in what the latest developments in representational learning and deep learning can contribute to this inquiry, and would like to explore these systems as part of agent-based installations to understand better how these notions interplay.

Finally, in terms of practice, there are a number of different artistic strategies that need to be further explored. Firstly, the idea of agent "shaping", where an agent is trained by the artist to give its behavior the shape he wants through direct feedback, is particularly worthy of attention.

One can imagine, for example, a robot that interacts in its environment; the artist observes it and can, through a remote control, give it positive or negative rewards depending on its actions, until the robot reaches a certain kind of behavior.

Secondly, as a strategy for reducing the time span taken by an agent to stabilize into a learned behavior, I would like to experiment with pre-training systems over a problem, saving the model's state regularly. This would generate a sequence of states representing the different stages of learning the agent has been through. For example, if the learning process takes a month, we could save one version of the weights every day, yielding thirty (30) sequential sets of parameters, going from the initial state to the fully adapted state. Once in the gallery space, this process could be accelerated by sequentially switching from one set of weights to the next, but clearly at a more rapid timescale. For instance, one could run through a sequence by loading a set of weights every 20 seconds, allowing the audience to see the adaptive procedure over a span of about 10 minutes.

Thirdly, I am interested in further exploring the idea of bidirectional coupling, the idea that an adaptive agent is different from an inanimate tool (such as a cane) because not only can it be "ready-to-hand" for a human user, it can also make the human user ready to its hand, so to speak. In that respect, I am interested in developing artistic agents that could live with a human for a while so that a symbiotic, bidirectional coupling could start to emerge between them. (I have already started working on this with a lamp project). In particular, I am interested in the relationships a human could develop with such a system, which could become an imprint of the human user's own behavior. The agent's adaptive features, when placed in an environment over a long time, could have some kind of historical capacity, not so much in terms of being a placeholder for content, such as pictures or words, but rather in its retention of a certain behavior and its aptitude to recognize. When interacting, both human and non-human agents would adapt to each other and, in the process, become attuned to one another. Thus, after a period of separation, when the two met again (provided that they did not change too much in the mean time) there could be a possibility that they might recognize one another. What becomes even more powerful is that the kind of interaction the prototype suggests offers a ground for an historical-based social orientation, where an object attuned to someone can interact with another person. The object is rather flexible in regards to to the kind of environment it falls into: it can as well be used solely by a single person, by two persons or by a group (eg. a family).

7.7 Final Thoughts

The advent of Machine Learning and its growing importance in the 21st century resonates with the performative turn in humanities, as it suggests that machinic intelligence has less to do with logic and rationality, and more with lifelike processes of self-organization that run beyond traditional frameworks of representation. While the general public still perceives computers as things that accumulate and manipulate data by applying logical rules, the artificial agents that will populate our future seem more likely to be akin to biological forms than to advanced calculators. However, the price to pay is that the behavior and data processing of these artificial “hybrids” might lay beyond their users’ comprehension — even more opaque than the computers of today — as these hypothetically adaptive devices would continuously come up with their own rules.

Artists have an important role to play in addressing these technologies beyond the applications that science and industry suggest. The autonomy and indeterminacy of adaptive systems suggest new ways to think about how to make art, how to experience art, and how the artistic world works. One strategy might be to go back to the experiments of early British cyberneticians, enacting a newly retro genre of performative science. In some ways, art might be more protected from neoliberal funding models that currently dominate scientific research, allowing artists to address blind spots left in the wake of an increasingly homogenic socio-economic scientific research culture. My ambition is that my own research presented in this thesis will allow artists and theorists to understand better the tensions between art and science in the field of AI, the role adaptation plays as a concept in these relationships, and show how artists can reemploy, exploit, and reappropriate techniques developed by science to create new aeshethic morphologies, just as Xenakis did decades ago when he used stochastics to generate new forms of music. I hope that this study, as well as my artistic practice, can actively participate to the development of this endeavor.

Bibliography

Abbeel, P., et al. 2007. “An Application of Reinforcement Learning to Aerobatic Helicopter Flight”. In *Advances in Neural Information Processing Systems 20*, 1–8.

Abbott, L. F. 1999. “Lapicque’s Introduction of the Integrate-and-Fire Model Neuron (1907)”. *Brain Research Bulletin* 50 (5–6): 303–304.

Alpaydin, E. 2004. *Introduction to Machine Learning.* Adaptive computation and machine learning. Cambridge, MA: MIT Press.

Anderson, J. A., and E. Rosenfeld. 1998. *Talking Nets an Oral History of Neural Networks.* Cambridge, MA: MIT Press.

Arel, I., D. Rose, and T. Karnowski. 2010. “Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier]”. *IEEE Computational Intelligence Magazine* 5 (4): 13–18.

Ascott, R. 2003a. “Behaviourist Art and the Cybernetic Vision”. In *Telematic Embrace: Visionary Theories of Art, Technology, and Consciousness*, 109–156. Berkeley: University of California Press.

— . 2003b. *Telematic Embrace: Visionary Theories of Art, Technology, and Consciousness.* Ed. by E. A. Shanken. Berkeley: University of California Press.

Ashby, W. R. 1957. *An Introduction to Cybernetics.* London: Chapman & Hall.

— . 1954. *Design for a Brain.* New York: Wiley.

Audry, S. 2010. “Absences: Public Art Interventions in Natural Spaces Using Autonomous Electronic Devices”. In *ISEA 2010 Conference Proceedings*, 469–471. Ruhr, Germany.

Austin, J. L. 1962. *How to Do Things with Words*. Vol. 88. London, UK: Oxford University Press.

Baginsky, N. A. 2005. “Aglaopheme - Slide Guitar Robot”. Visited on 03/08/2014. `http://www.baginsky.de/agl/agl_index.html`.

Baljko, M., and N. Tenhaaf. 2008. “The Aesthetics of Emergence: Co-Constructed Interactions”. *ACM Transactions on Computer-Human Interaction* 15 (3): 11:1–11:27.

Bates, J. 1994. “The Role of Emotion in Believable Agents”. *Commun. ACM* 37 (7): 122–125.

Bedau, M. A. 2000. “Artificial Life VII: Looking Backward, Looking Forward (Editor’s Introduction to the Special Issue)”. *Artif. Life* 6 (4): 261–264.

Bedau, M. A., and C. E. Cleland. 2010. *The Nature of Life: Classical and Contemporary Perspectives from Philosophy and Science*. 1st ed. Cambridge University Press.

Bengio, Y., A. Courville, and P. Vincent. 2013. “Representation Learning: A Review and New Perspectives”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (8): 1798–1828.

Bengio, Y. 2009. “Learning Deep Architectures for AI”. *Foundations and Trends in Machine Learning* 2 (1): 1–127.

Bengio, Y., and J.-S. Senécal. 2003. “Quick Training of Probabilistic Neural Nets by Importance Sampling”. In *Proceedings of the Conference on Artificial Intelligence and Statistics (AISTATS)*.

Bengio, Y., et al. 2006. “Neural Probabilistic Language Models”. In *Innovations in Machine Learning*, ed. by P. D. E. Holmes and P. L. C. Jain, 137–186. Studies in Fuzziness and Soft Computing 194. Springer Berlin Heidelberg.

Bernard, C. 1957. *An Introduction to the Study of Experimental Medicine*. New York: Dover Publications.

Bertalanffy, L. van. 1969. *General System Theory: Foundations, Development, Applications*. New York: Braziller.

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition.* Oxford University Press.

Boden, M. A. 1996. "Autonomy and Artificiality". In *The Philosophy of Artificial Life*, First, ed. by M. A. Boden, 95–108. Oxford University Press, USA.

— . 2009. "Computer Models of Creativity". *AI Magazine* 30 (3): 23.

— . 2006. *Mind as Machine: A History of Cognitive Science.* Oxford : New York: Clarendon Press ; Oxford University Press.

Boden, M. A. 2004. *The Creative Mind Myths and Mechanisms.* London; New York: Routledge.

Bogart, B. D. R., and P. Pasquier. 2013. "Context Machines: A Series of Situated and Self-Organizing Artworks". *Leonardo* 46 (2): 114–143. JSTOR: 23468147.

Boulanger-lewandowski, N., Y. Bengio, and P. Vincent. 2012. "Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription". In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, ed. by J. Langford and J. Pineau, 1159–1166. New York, NY, USA: ACM.

Bourgine, P., and F. J. Varela. 1994. "Towards a Practice of Autonomous Systems". In *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, xi–xvii. Paris, France: The MIT Press.

Bown, O. 2012. "Generative and Adaptive Creativity: A Unified Approach to Creativity in Nature, Humans and Machines (Preprint)". In *Computers and Creativity*, ed. by J. McCormack and M. d'Inverno, 361–381. Springer Berlin Heidelberg.

Breazeal, C. 2002. *Designing Sociable Robots.* Cambridge, MA: The MIT Press.

Brooks, R. 1986. "A Robust Layered Control System for a Mobile Robot". *IEEE Journal on Robotics and Automation* 2 (1): 14–23.

Brooks, R. 2002. *Flesh and Machines: How Robots Will Change Us.* Random House LLC.

Brooks, R. A. 1989. *A Robot that Walks: Emergent Behaviors from a Carefully Evolved Network* AI MEMO 1091. Massachusets Institute of Technology.

— . 1987. "Intelligence Without Representation". *Artificial Intelligence* 47:139–159.

Brooks, R. A. 1999. *Cambrian Intelligence: The Early History of the New AI.* Cambridge, MA: MIT Press.

Brown, R., et al. 2001. *Biotica: Art, Emergence and Artificial Life.* London: RCA CRD Research.

Burnham, J. 1969. "Real Time Systems". *Artforum* 8 (1): 49–55.

— . 1968. "Systems Esthetics". *Artforum* 7 (1): 30–35.

— . 1970. "The Aesthetics of Intelligent Systems". In *On the Future of Art*, ed. by E. F. Fry, 95–122. New York: Viking Press.

Busch, K. 2011. "Artistic Research and the Poetics of Knowledge". *Art & Research: A Journal of Ideas, Contexts and Methods* 2 (2).

Butler, J. 1999. *Gender Trouble: Feminism and the Subversion of Identity.* New York: Routledge.

— . 2004. *Undoing Gender.* New York; London: Routledge.

Cage, J. 1961. "Experimental Music". In *Silence*, 7–12. Middletown, CT: Wesleyan University Press.

Cage, J., and D. Charles. 1981. *For the Birds.* London, UK: Marion Boyars Publishers Ltd (.

Callon, M. 2006. *What Does It Mean to Say that Economics Is Performative?* CSI WORKING PAPERS SERIES 005. Centre de Sociologie de l'Innovation (CSI), Mines ParisTech.

Campbell, S. R., D. L. Wang, and C. Jayaprakash. 1999. "Synchrony and Desynchrony in Integrate-and-Fire Oscillators". *Neural Computation* 11 (7): 1595–1619.

Cannon, W. B. 1928. "The Mechanism of Emotional Disturbance of Bodily Functions". *New England Journal of Medicine* 198 (17): 877–884.

Cariani, P. 1990. "Adaptive Connection to the World through Self-Organizing Sensors and Effectors". In *Proceedings. 5th IEEE International Symposium on Intelligent Control 1990*, 1:73–78. IEEE.

— . 1991. "Adaptivity and Emergence in Organisms and Devices". *World Futures* 32 (2-3): 111–132.

— . 2012. "Creating New Informational Primitives in Minds and Machines". In *Computers and Creativity*, ed. by J. McCormack and M. d'Inverno, 383–417. Springer Berlin Heidelberg.

Cariani, P. A. 2016. “Beware False Dichotomies”. *Constructivist Foundations* 11 (3): 472–475.

— . 2008. “Emergence and Creativity”. In *Emoção Art.ficial 4.0: Emergência*, 20–42. Sao Paolo, Brazil: Itaú Cultural.

— . 1989. “On the Design of Devices with Emergent Semantic Functions”, State University of New York at Binghamton.

Chalmers, D. J. 1997. *The Conscious Mind: In Search of a Fundamental Theory.* Philosophy of Mind Series. Oxford University Press.

Chalup, S., C. Murch, and M. Quinlan. 2007. “Machine Learning With AIBO Robots in the Four-Legged League of RoboCup”. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 37 (3): 297–310.

Church, A. 1936. “A Note on the Entscheidungsproblem”. *J. Symb. Log.* 1 (1): 40–41.

Clarke, B., and M. B. N. Hansen, eds. 2009. *Emergence and Embodiment: New Essays on Second-Order Systems Theory.* Duke University Press Books.

Cohen, A. E., and F. J. Varela. 2000. “Facing Up to the Embarrassment: The Practice of Subjectivity in Neuroscientific and Psychoanalytic Experience”. *Journal of European Psychoanalysis* (10-11).

Cooper, S. J. 2005. “Donald O. Hebb’s Synapse and Learning Rule: A History and Commentary”. *Neuroscience & Biobehavioral Reviews* 28 (8): 851–874.

Cybernetics Serendipity - Late Night Lineup. 1968. In collab. with J. Reichardt.

Dağdelen, B., and Y. E. Kucur. 2016. “Bio-Inspire FullDome Performance”. Visited on 09/26/2016. `http://prix2016.aec.at/prixwinner/17912/`.

D’Avila Garcez, A. S., K. Broda, and D. M. Gabbay. 2001. “Symbolic Knowledge Extraction from Trained Neural Networks: A Sound Approach”. *Artificial Intelligence* 125 (1): 155–207.

Dawkins, R. 1986. *The Blind Watchmaker.* 1st American ed. New York: Norton.

Demers, L.-P., and B. Vorn. 1995. “Real Artificial Life as an Immersive Media”. In *Convergence: Proceedings of the 5th Biennial Symposium for Arts and Technology*, 190–203. New London, CT: Center for Arts and Technology at Connecticut College.

Dewey, J. 1959. *Art as Experience*. A Perigee book. New York, NY: Putnam.

Di Scipio, A. 1994. “Formal Processes of Timbre Composition: Challenging the Dualistic Paradigm of Computer Music”. In *Proceedings of the 1994 International Computer Music Conference*, 202–208. San Francisco: : International Computer Music Association.

Dietrich, E. 1990. “Computationalism”. *Social Epistemology* 4 (2): 135–154.

Dinneen, G. P. 1955. “Programming Pattern Recognition”. In *Proceedings of the March 1-3, 1955, Western Joint Computer Conference*, 94–100. AFIPS ’55 (Western). New York, NY, USA: ACM.

Dorigo, M., and M. Colombetti. 1997. *Robot Shaping: An Experiment in Behavior Engineering*. Cambridge, Mass: A Bradford Book.

Dourish, P. 2001. *Where The Action Is: The Foundations of Embodied Interaction*. Kindle Edition. The MIT Press.

Downie, M. 2005. “Choreographing the Extended Agent: Performance Graphics for Dance Theater”. Ph. D. Thesis, Massachusetts Institute of Technology. Visited on 03/12/2013.

Dreyfus, H. L. 1979. *What Computers Can’t Do: The Limits of Artificial Intelligence*. Harper & Row.

Duda, R. O., P. E. Hart, and D. G. Stork. 2001. *Pattern Classification*. 2nd ed. New York: Wiley.

Dupuy, J. P. 2000. *The Mechanization of the Mind: On the Origins of Cognitive Science*. New French thought. Princeton: Princeton University Press.

Eck, D. 2016. “Welcome to Magenta!” Visited on 08/04/2016. `https://magenta.tensorflow.org/welcome-to-magenta`.

Eck, D., and J. Schmidhuber. 2002. *A First Look at Music Composition Using LSTM Recurrent Neural Networks*. Technical Report. Manno, Switzerland: Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale.

Edmonds, E., Z. Bilda, and L. Muller. 2009. “Artist, Evaluator and Curator: Three Viewpoints on Interactive Art, Evaluation and Audience Experience”. *Digital Creativity* 20 (3): 141–151.

Edwards, R., and J. Holland. 2013. *What Is Qualitative Interviewing?* A&C Black.

Eigenfeldt, A., A. Burnett, and P. Pasquier. 2012. “Evaluating Musical Metacreation in a Live Performance Context”. *Proceedings of the Third International Conference on Computational Creativity*: 140–144.

Fernández, M. 2006. “‘Life-Like’: Historicizing Process and Responsiveness in Digital Art”. In *The Art of Art History: A Critical Anthology*, New ed, ed. by D. Preziosi, 477–487. Oxford history of art. Oxford ; New York: Oxford University Press.

Fischer-Lichte, E. 2008. *The Transformative Power of Performance: Re-Enchanting the World*. London: Routledge.

Fogel, L. J., A. J. Owens, and M. J. Walsh. 1967. *Artificial Intelligence Through Simulated Evolution*. New York, NY: John Wiley and Sons.

Freeman, W. J. 2000. *How Brains Make up their Minds*. Maps of the mind. New York: Columbia University Press.

— . 1995. *Societies of Brains: A Study in the Neuroscience of Love and Hate*. The International Neural Networks Society series. Hillsdale, NJ: Lawrence Erlbaum Associates.

Friedman, T. L. 2012. “Come the Revolution”. *The New York Times*.

Gemeinboeck, P., and R. Saunders. 2013. “Creative Machine Performance: Computational Creativity and Robotic Art”. In *Proceedings of the Fourth International Conference on Computational Creativity*. ICCC2013. Sydney: International Association for Computational Creativity.

Gleick, J. 2008. *Chaos: Making a New Science*. Revised. Penguin Books.

Glynn, R. 2008. “Conversational Environments Revisited”. In *Conference Proceedings for the 19th European Meeting on Cybernetics and Systems Research*. Vienna, Austria.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. “Deep Learning”.

Grefenstette, J. J., D. E. Moriarty, and A. C. Schultz. 2011. “Evolutionary Algorithms for Reinforcement Learning”. arXiv: `1106.0221 [cs]`.

Gu, D., and H. Hu. 2002. “Reinforcement Learning of Fuzzy Logic Controllers for Quadruped Walking Robots”. In *15th Triennial World Congress*. Barcelona, Spain.

Halpern, O. 2014. *Beautiful Data: A History of Vision and Reason since 1945*. Experimental futures. Durham: Duke University Press.

Harnad, S. 2000. "Correlation vs. Causality: How/Why the Mind/Body Problem Is Hard". *Journal of Consciousness Studies* 7 (4): 54–61.

— . 2007. "Maturana's Autopoietic Hermeneutics versus Turing's Causal Methodology for Explaining Cognition". *Pragmatics & Cognition* 15 (3): 599–603.

— . 1991. "Other Bodies, Other Minds: A Machine Incarnation of an Old Philosophical Problem". *Minds and Machines* 1 (1): 43–54.

— . 2008. "The Annotation Game: On Turing (1950) on Computing, Machinery and Intelligence". In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, ed. by G. Epstein , Robert & Peters. Springer.

— . 1990. "The Symbol Grounding Problem". *Physica D: Nonlinear Phenomena* 42 (1-3): 335–346.

— . 2005. "To Cognize Is to Categorize: Cognition Is Categorization". In *Handbook of Categorization. Summer Institute in Cognitive Sciences on Categorisation*, ed. by C. Lefebvre and H. Cohen. Elsevier.

— . 2001. "What's Wrong and Right About Searle's Chinese Room Argument?" In *Essays on Searle's Chinese Room Argument*, ed. by J. Preston and M. Bishop. Oxford University Press.

Hayles, N. K. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press.

Hebb, D. O. 1949. *The Organization of Behavior*. New York: Wiley & Sons.

Heidegger, M. 1972. *On Time and Being*. New York: Harper & Row.

Helmreich, S. 2000. *Silicon Second Nature: Culturing Artificial Life in a Digital World*. Updated. Berkeley; Los Angeles: University of California Press.

Heylighen, F. 2000. "Foundations and Methodology for an Evolutionary World View: A Review of the Principia Cybernetica Project". *Foundations of Science* 5 (4): 457–490.

Hinton, G. E., S. Osindero, and Y. W. Teh. 2006. "A Fast Learning Algorithm for Deep Belief Nets". *Neural Computation* 18 (7): 1527–1554.

Holland, J. H. 1992. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. New edition edition. Cambridge, MA: A Bradford Book.

— . 1996. *Hidden Order: How Adaptation Builds Complexity.* Redwood City, CA: Addison Wesley Longman Publishing Co. Inc.

Hoyt, S., S. Mckennoch, and L. G. Bushnell. 2005. *An Autonomous Multi-Agent Testbed Using Infrared Wireless Communication and Localization.* Technical Report UWEETR-2005-3/17/05. Seattle, Washington: Department of Electrical Engineering, University of Washington.

Hultén, P., ed. 1968. *The Machine As Seen at the End of the Mechanical Age.* New York, NY: Museum of Modern Art.

Idel, M. 1990. *Golem: Jewish Magical and Mystical Traditions on the Artificial Anthropoid.* SUNY series in Judaica. Albany, NY: State University of New York Press.

Isla, D. 2005. "Handling Complexity in the Halo 2 AI". In *Game Developer's Conference 2005 Proceedings.* San Francisco, CA: CMP Inc.

Johnson, C. G., and J. J. R. Cardalda. 2002. "Introduction: Genetic Algorithms in Visual Art and Music". *Leonardo* 35 (2): 175–184. JSTOR: `1577199`.

Johnston, J. 2008. *The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI.* Cambridge, MA: MIT Press.

Kac, E. 1997. "Foundation and Development of Robotic Art". *Art Journal* 56:60–67.

Kelly, S. 2016. "Stephen Kelly: Open Ended Ensemble (Competitive Coevolution)". Visited on 07/19/2016. `http://www.theinc.ca/exhibitions/stephen-kelly/`.

Kember, S. 2002. *Cyberfeminism and Artificial Life.* 1st ed. London; New York: Routledge.

Kennedy, J. F., R. C. Eberhart, and Y. Shi. 2001. *Swarm Intelligence.* San Francisco, CA: Morgan Kaufmann.

Kim, J. F., and K. Galvin. 2012. “An Interview with Simon Penny: Techno-Utopianism, Embodied Interaction and the Aesthetics of Behavior”. Ed. by L. Aceti and S. Penny. *Leonardo Electronic Almanac (DAC09: After Media: Embodiment and Context)* 17 (2): 136–145.

Klein, Y. A. 1998. “Living Sculpture: The Art and Science of Creating Robotic Life”. *Leonardo* 31 (5): 393.

— . 2014. “Yves Amu Klein Living Sculpture”. `http://www.livingsculpture.com/`.

Klein, Y. A., and M. Hudson. 2003. “Living Sculpture”. In *Proceedings of the Fifth International Conference on Ubiquitous Computing*.

Kogan, G., and F. Tseng. 2016. “Machine Learning for Artists”. Visited on 09/26/2016. `https://ml4a.github.io/`.

Kohonen, T. 1981. “Automatic Formation of Topological Maps of Patterns in a Self-Organizing System”. In *Proceedings of 2SCIA*, ed. by E. Oja and O. Simula, 214–220. Helsinki, Finland.

— . 2001. *Self-Organizing Maps*. 3rd ed. Springer series in information sciences 30. Berlin ; New York: Springer.

Kornienko, S., and S. Kornienko. 2011. “IR-Based Communication and Perception in Microrobotic Swarms”. arXiv: `1109.3617 [cs]`.

Kruger, N., et al. 2013. “Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision?” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (8): 1847–1871.

Kurzweil, R. 2006. *The Singularity Is Near: When Humans Transcend Biology*. New York: Penguin Books.

Kwastek, K. 2013. *Aesthetics of Interaction in Digital Art*. Cambridge, MA: The MIT Press.

Lange, S., T. Gabel, and M. Riedmiller. 2011. “Batch Reinforcement Learning”. In *Reinforcement Learning: State of the Art*, ed. by M. Wiering and M. van Otterlo, 1–28. Heidelberg; New York: Springer.

Langley, P. 2011. “The Changing Science of Machine Learning”. *Machine Learning* 82 (3): 275–279.

Langton, C. G. 1990. “Computation at the Edge of Chaos: Phase Transitions and Emergent Computation”. *Physica D: Nonlinear Phenomena* 42 (1–3): 12–37.

Langton, C. G. 1989a. “Artificial Life”. In *Artificial Life*, ed. by C. G. Langton, 1–47. Redwood City, CA: Addison-Wesley.

— , ed. 1995. *Artificial Life: An Overview.* Cambridge, MA: The MIT Press.

— , ed. 1989b. *Artificial Life: Proceedings Of An Interdisciplinary Workshop On The Synthesis And Simulation Of Living Systems.* Vol. 6. Santa Fe Institute studies in the sciences of complexity. Redwood City, Calif: Addison-Wesley.

— . 1984. “Self-Reproduction in Cellular Automata”. *Physica D: Nonlinear Phenomena* 10 (1–2): 135–144.

— . 1986. “Studying Artificial Life with Cellular Automata”. *Proceedings of the Fifth Annual International Conference* 22 (1–3): 120–149.

Latour, B. 1991. *Nous N’avons Jamais Été Modernes. Essai D’anthropologie Symétrique.* Paris: La Découverte.

— . 2005. *Reassembling the Social: An Introduction to Actor-Network Theory.* Oxford & New York: Oxford University Press.

Law, E. L.-C., et al. 2009. “Understanding, Scoping and Defining User Experience: A Survey Approach”. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 719–728. CHI ’09. New York, NY: ACM.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. “Deep Learning”. *Nature* 521 (7553): 436–444.

Lewis, C. 1982. *Using The "thinking Aloud" Method in Cognitive Interface Design.* Technical Report RC-9265. IBM T.J. Watson Research Center.

Maass, W., and C. M. Bishop. 2001. *Pulsed Neural Networks.* MIT Press.

Madrigal, A. C. 2015. “The Case against Killer Robots, from a Guy Actually Working on Artificial Intelligence”. *Fusion.*

Maes, P. 1994. “Modeling Adaptive Autonomous Agents”. *Artificial Life* 1:135–162.

Maes, P., et al. 1990. “Learning to Coordinate Behaviors”. In *Association for the Advancement of Artificial Intelligence*, 2:796–802. Palo Alto, CA: AAAI Press.

Markoff, J. 2013. “Brainlike Computers, Learning From Experience”. *The New York Times*.

Marzinotto, A., et al. 2014. “Towards a Unified Behavior Trees Framework for Robot Control”. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 5420–5427.

Mateas, M. 2001. “Expressive AI: A Hybrid Art and Science Practice”. *Leonardo* 34 (2): 147–153. JSTOR: `1577018`.

Mathews, S. 2005. “The Fun Palace: Cedric Price’s Experiment in Architecture and Technology”. *Technoetic Arts: A Journal of Speculative Research* 3 (2): 73–91.

Maturana, H. R., and F. J. Varela. 1980. *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: D. Reidel.

Maxwell, J. C. 1867. “On Governors”. *Proceedings of the Royal Society of London* 16:270–283. JSTOR: `112510`.

McCarthy, J., et al. 2006. “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955”. *AI Magazine* 27 (4): 12.

McCormack, J. 2009. “The Evolution of Sonic Ecosystems”. In *Artificial Life Models in Software*, 2. ed, ed. by M. Komosinski and A. Adamatzky, 393–414. London: Springer.

McCulloch, W. S., and W. Pitts. 1943. “A Logical Calculus of the Ideas Immanent in Nervous Activity”. *The bulletin of mathematical biophysics* 5 (4): 115–133.

McElhearn, K. 2016. “John Cage and the Anechoic Chamber”. Visited on 07/26/2016. `http://www.mcelhearn.com/john-cage-and-the-anechoic-chamber/`.

McLurkin, J. 2004. “Stupid Robot Tricks: A Behavior-Based Distributed Algorithm Library for Programming Swarms of Robots”. S.M. Thesis, Massachusetts Institute of Technology.

Medler, D. A. 1998. “A Brief History of Connectionism”. *Neural Computing Surveys* 1:61–101.

Miller, A. I. 2014. *Colliding Worlds: How Cutting-Edge Science Is Redefining Contemporary Art*. New York: WW Norton.

Minsky, M. L., and S. Papert. 1969. *Perceptrons: An Introduction to Computational Geometry.* Cambridge, MA: MIT Press.

Mitchell, M. 1998. *An Introduction to Genetic Algorithms.* Third Printing. Cambridge, MA: The MIT Press.

— . 1995. “Genetic Algorithms: An Overview”. *Complexity* 1 (1): 31–39.

Mnih, V., et al. 2015. “Human-Level Control through Deep Reinforcement Learning”. *Nature* 518 (7540): 529–533.

Mnih, V., et al. 2013. “Playing Atari with Deep Reinforcement Learning”. In *NIPS 2013 Deep Learning Workshop.*

Mordvintsev, A., C. Olah, and M. Tyka. 2015. “Inceptionism: Going Deeper into Neural Networks”. Visited on 06/03/2016. `http://googleresearch.blogspot.com/2015/06/inceptionism-going-deeper-into-neural.html`.

Nath, V., and S. E. Levinson. 2014. *Autonomous Robotics and Deep Learning.* 2014 edition. Springer.

Newell, A. 1955. “The Chess Machine: An Example of Dealing with a Complex Task by Adaptation”. In *Proceedings of the March 1-3, 1955, Western Joint Computer Conference*, 101–108. AFIPS ’55 (Western). New York, NY, USA: ACM.

Newell, A., J. C. Shaw, and H. A. Simon. 1958. “Elements of a Theory of Human Problem Solving.” *Psychological Review* 65 (3): 151.

Newell, A., J. C. Shaw, and H. A. Simon. 1959. “Report on a General Problem-Solving Program”. In *Proceedings of the International Conference on Information Processing*, 256–264.

Ng, A. Y., et al. 2004. “Autonomous Helicopter Flight via Reinforcement Learning”. In *Advances in Neural Information Processing Systems 16 (NIPS 2003)*, 799–806.

Nilsson, N. J. 2010. *The Quest for Artificial Intelligence: A History of Ideas and Achievements.* Cambridge, UK: Cambridge University Press.

Ovid. 2008. *Metamorphoses.* In collab. with E. J. Kenney. Trans. by A. D. Melville. Oxford: Oxford University Press.

Paparrigopoulos, K. 2011. “Divergences and Convergences between Xenakis and Cage’s Indeterminism”. In *Proceedings of the Xenakis International Symposium*. London.

Pask, G. 1971. “A Comment, a Case History and a Plan”. In *Cybernetics, Art, and Ideas.* Ed. by J. Reichardt, 76–99. Greenwich, CN: New York Graphic Society.

— . 1968. *An Approach to Cybernetics*. London: Hutchinson.

Pasquier, P. 2015. “Metacreation Lab: Generative Systems”. Simon Fraser University.

Penny, S. 2000. “Agents as Artworks and Agent Design as Artistic Practice”. In *Advances in Consciousness Research*, ed. by K. Dautenhahn, 19:395–414. Amsterdam: John Benjamins Publishing Company.

— . 2009. “Art and Artificial Life – A Primer”. In *Proceedings of the Digital Arts and Culture Conference*. Irvine: University of California.

— . 2008. “Bridging Two Cultures – towards a History of the Artist-Inventor”. In *Artists as Inventors, Inventors as Artists, Anthology of Ludwig Boltzmann Institute*, ed. by D. Daniels and B. U. Schmidt. Austria: Hatje Cantz.

— . 1997. “Embodied Cultural Agents: At the Intersection of Robotics, Cognitive Science and Interactive Art”. In *AAAI Socially Intelligent Agents: Papers from the 1997 Fall Symposium*, ed. by K. Dautenhahn, 103–105. Menlo Park: AAAI Press.

Pfeifer, R. 1996. “Symbols, Patterns, and Behavior: Towards a New Understanding of Intelligence”. In *Proc. Japanese Conference on Artificial Intelligence*, 1–15.

Picard, R. W. 2000. *Affective Computing*. Cambridge, MA: MIT Press.

Pickering, A. 2013. “Being in an Environment: A Performative Perspective”. *Natures Sciences Sociétés* 21 (1): 77–83.

— . 2010. *The Cybernetic Brain: Sketches of Another Future*. Chicago; London: University of Chicago Press.

— . 1995. *The Mangle of Practice: Time, Agency, and Science*. 1st ed. Chicago: University Of Chicago Press.

— . 2009. “The Politics of Theory: Producing Another World, with Some Thoughts on Latour”.

Quinlan, M. 2006. “Machine Learning on AIBO Robots”. Ph. D. Thesis, University of Newcastle.

Rasmusson, J. 2010. *The Agile Samurai: How Agile Masters Deliver Great Software*. The pragmatic programmers. Raleigh: The Pragmatic Bookshelf.

Ray, T. S. 1991. “An Approach to the Synthesis of Life”. In *Artificial Life II*, ed. by C. G. Langton et al., 10:371–408. Redwood City, CA: Addison-Wesley.

Rechenberg, I. 1965/August//. “Cybernetic Solution Path of an Experimental Problem”. In *Royal Aircraft Establishment Translation No. 1122, B. F. Toms, Trans.* Ministry of Aviation, Royal Aircraft Establishment.

Rechenberg, I. 1973. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Reichardt, J., ed. 1968. *Cybernetic Serendipity: The Computer and the Arts*. 2nd ed. New York, NY: Studio International.

Reynolds, C. W. 1987. “Flocks, Herds, and Schools: A Distributed Behavioral Model”. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, 25–34. SIGGRAPH ’87. New York, NY.

Riedmiller, M., and A. Merke. 2002. “Using Machine Learning Techniques in Complex Multi-Agent Domains”. In *Adaptivity and Learning*. Springer.

Rinaldo, K. E. 1998. “Technology Recapitulates Phylogeny: Artificial Life Art”. *Leonardo* 31 (5): pp. 371–376.

Rosenblatt, F. 1957. *The Perceptron – a Perceiving and Recognizing Automaton* 85-460-1. Buffalo, NY: Cornell Aeronautical Laboratory.

Rosenblueth, A., N. Wiener, and J. Bigelow. 1943. “Behavior, Purpose and Teleology”. *Philosophy of Science* 10 (1): pp. 18–24.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. “Learning Representations by Back-Propagating Errors”. *Nature* 323:533–536.

Samuel, A. L. 1959. “Some Studies in Machine Learning Using the Game of Checkers”. *IBM Journal of Research and Development* 3 (3): 210–229.

Schechner, R. 2003. *Performance Theory.* London & New York: Routledge.

Schmidhuber, J. 2015. “Deep Learning in Neural Networks: An Overview”. *Neural Networks* 61:85–117.

Schmidhuber, J., et al. 2011. “On Fast Deep Nets for AGI Vision”. In *Artificial General Intelligence*, ed. by J. Schmidhuber, K. R. Thórisson, and M. Looks, 243–246. Lecture Notes in Computer Science 6830. Springer Berlin Heidelberg.

Schöffer, N. 2004. *Nicolas Schöffer.* In collab. with Villa Tamaris (La Seyne-sur-Mer, France) and Fondation Vasarely. Art contemporain 9. Dijon: Presses du réel.

Searle, J. R. 1980. “Minds, Brains, and Programs”. *Behavioral and Brain Sciences* 3:417–424.

Sedgwick, E. K. 1993. “Queer Performativity: Henry James’s The Art of the Novel”. *GLQ: a journal of Lesbian and Gay Studies* 1 (1): 1–16.

Selfridge, O. G. 1955. “Pattern Recognition and Modern Computers”. In *Proceedings of the March 1-3, 1955, Western Joint Computer Conference*, 91–93. AFIPS ’55 (Western). New York, NY, USA: ACM.

Selfridge, O. G. 1959. “Pandemonium: A Paradigm for Learning”. In *Symposium on the Mechanization of Thought Processes*, ed. by D. K. Blake and A. M. Uttley, 1:511–531. London: HMSO.

Senécal, J.-S. 2003. “Accélérer L’entraînement D’un Modèle Non-Paramétrique de Densité Non Normalisée Par Échantillonnage Aléatoire”. Master Thesis, Université de Montréal.

Serre, T., et al. 2007. “A Quantitative Theory of Immediate Visual Recognition”. *Progress in brain research* 165. pmid: 17925239.

Serres, M. 1982. *The Parasite.* In collab. with L. R. Schehr. Baltimore: Johns Hopkins University Press.

Shanken, E. A. 2002. “Cybernetics and Art: Cultural Convergence in the 1960s”. In *From Energy to Information*, ed. by B. Clarke and L. Darlymple, 155–177. Palo Alto, NM: Stanford University Press.

— , ed. 2015. *Systems*. Cambridge, MA: The MIT Press.

Shannon, C. E. 1948. “A Mathematical Theory of Communication”. *Bell System Technical Journal* 27 (3): 379–423.

Shiffman, D. 2012. *The Nature of Code*. Ed. by S. Fry.

Sims, K. 1991. “Artificial Evolution for Computer Graphics”. In *Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques*, 319–328. SIGGRAPH ’91. New York, NY: ACM.

— . 1994. “Evolving Virtual Creatures”. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, 15–22. SIGGRAPH ’94. New York, NY: ACM.

Skarda, C. A. 1992. “Perception, Connectionism, and Cognitive Science”. In *Understanding Origins*, ed. by F. J. Varela and J.-P. Dupuy, 265–271. Dordrecht: Springer Netherlands.

Smith, A. W. 1974. “Information Theory and Cybernetics”. *Journal of Cybernetics* 4 (3): 1–5.

Software: Information Technology: It’s New Meaning for Art. 1970. In collab. with J. Burnham. New York, NY: Jewish Museum.

Soler-Adillon, J. 2015. “Emergence as Self-Organization and as Generation of Novelty: A Framework for Understanding Emergence in the Context of Interactive Art”, Universitat Pompeu Fabra.

Solomos, M. 2006. “Cellular automata in Xenakis’s music. Theory and Practice”. In *Definitive Proceedings of the International Symposium Iannis Xenakis*, ed. by S. Makis, A. Georgaki, and G. Zervos, 11 p. Greece.

Soni, V., and S. Singh. 2006. “Reinforcement Learning of Hierarchical Skills on the Sony Aibo Robot”. In *Proceedings of the 5th International Conference on Development and Learning*.

Suchman, L. A. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge University Press.

Sutton, R. S., and A. G. Barto. 1998. *Reinforcement Learning: An Introduction.* Adaptive computation and machine learning. Cambridge, MA: MIT Press.

Tenhaaf, N. 2008. "Art Embodies A-Life: The Vida Competition". *Leonardo* 41 (1): 6–15.

— . 2000. "Perceptions of Self in Art and Intelligent Agents". In *Proceedings of the 2000 AAAI Fall Symposium on Socially Intelligent Agents: The Human in the Loop*, ed. by K. Dautenhahn. North Falmouth: AAAI Press.

Thivierge, J.-P., and P. Cisek. 2010. "Spiking Neurons that Keep the Rhythm". *Journal of Computational Neuroscience* 30 (3): 589–605.

Thompson, A. 1996. "An Evolved Circuit, Intrinsic in Silicon, Entwined with Physics". In *Proceedings of the First International Conference on Evolvable Systems: From Biology to Hardware*, 390–405. ICES '96. London, UK, UK: Springer-Verlag.

Tomlinson, B., and B. Blumberg. 2002. "AlphaWolf: Social Learning, Emotion and Development in Autonomous Virtual Agents". In *In Proceedings of First GSFC/JPL Workshop on Radical Agent Concepts*, 35–45.

Turing, A. M. 1950. "Computing Machinery and Intelligence". *Mind*, New Series, 59 (236): pp. 433–460.

— . 1936. "On Computable Numbers, with an Application to the Entscheidungsproblem". In *Proceedings of the London Mathematical Society*, 42:230–265. 2.

Turkle, S. 2006. *A Nascent Robotics Culture: New Complicities for Companionship*, AAAI Technical Report Series.

— . 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other.* New York: Basic Books.

— . 2007. *Evocative Objects: Things We Think With.* The MIT Press.

Turner, V. W. 1982. *From Ritual to Theatre: The Human Seriousness of Play.* Performance studies series, 1st v. New York City: Performing Arts Journal Publications.

Tyrrell, A., G. Auer, and C. Bettstetter. 2006. "Fireflies As Role Models for Synchronization in Ad Hoc Networks". In *Proceedings of the 1st International Conference on Bio Inspired Models of Network, Information and Computing Systems*. BIONETICS '06. New York, NY, USA: ACM.

Urbanowicz, R. J., and J. H. Moore. 2009. "Learning Classifier Systems: A Complete Introduction, Review, and Roadmap". *Journal of Artificial Evolution & Applications* 2009:1–25.

Van Heijenoort, J. 1977. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Cambridge: Harvard University Press.

Vapnik, V. N. 2000. *The Nature of Statistical Learning Theory*. 2nd ed. Statistics for engineering and information science. New York: Springer.

Varela, F. J. 1992. "Autopoiesis and a Biology of Intentionality". In *Proceedings of the Workshop Autopoiesis and Perception*, ed. by B. McMullin, 4–14. Dublin.

Varela, F. J., and P. Bourgine, eds. 1992. *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*. Cambridge, MA: MIT Press.

Varela, F. J., and J. Shear. 1999. "First-Person Methodologies: What, Why, How?" *Journal of Consciousness Studies* 6 (2-3): 1–14.

Varela, F. J., E. Thompson, and E. Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Von Foerster, H. 2003. *Understanding Understanding: Essays on Cybernetics and Cognition*. New York: Springer.

Von Neumann, J. 1951. "The General and Logical Theory of Automata". In *Cerebral Mechanisms in Behavior: The Hixon Symposium*, ed. by L. Jeffress, 1–41. John Wiley and Sons.

— . 1966. *Theory of Self-Reproducing Automata*. Ed. by A. W. Burks. Champaign, IL, USA: University of Illinois Press.

Von Uexküll, J. 1957. "A Stroll Through the Worlds of Animals and Men". In *Instinctive Behavior: The Development of Modern Concept*, ed. by C. H. Schiller, 5–80. Intl Universities Pr Inc.

Walter, W. G. 1951. "A Machine that Learns". *Scientific American* 185 (2): 60–63.

— . 1950. “An Imitation of Life”. *Scientific American* 182 (5): 42–45.

Weitz, M. 1956. “The Role of Theory in Aesthetics”. *The Journal of Aesthetics and Art Criticism* 15 (1): 27–35. JSTOR: `427491`.

Weizenbaum, J. 1976. *Computer Power and Human Reason: From Judgment to Calculation.* New York, NY, USA: W. H. Freeman & Co.

Wells, H. G. 1995. *The Time Machine*. New York: Dover Publications.

Whitelaw, M. Wednesday, August 16, 2006. “Jonathan McCabe - Nervous States”. Visited on 06/03/2016. `http://teemingvoid.blogspot.com.au/2006/08/jonathan-mccabe-nervous-states.html`.

— . 2004. *Metacreation: Art and Artificial Life*. Cambridge, MA: The MIT Press.

Widrow, B., and M. E. Hoff. 1960. “Adaptive Switching Circuits”. *1960 IRE WESCON Convention Record*: 96–104.

Wiener, N. 1961. *Cybernetics: Or Control and Communication in the Animal and the Machine.* Cambridge, MA: MIT Press.

Wiering, M., and M. Otterlo, eds. 2012. *Reinforcement Learning: State-of-the-Art.* 1st ed. ALO 12. New York: Springer.

Wilk, E. 2016. “The Artist-in-Consultance: Welcome to the New Management | E-Flux”. Visited on 08/04/2016. `http://www.e-flux.com/journal/the-artist-in-consultance-welcome-to-the-new-management/`.

Winograd, T. 1970. “Procedures as a Representation for Data in a Computer Program for Understanding Natural Language”. Thesis, Massachusetts Institute of Technology. Visited on 10/01/2016.

Winograd, T., and F. Flores. 1987. *Understanding Computers and Cognition: A New Foundation for Design.* Boston, MA: Addison-Wesley.

Wolfram, S. 2002. *A New Kind of Science*. Champaign, IL: Wolfram Media.

— . 1986. “Random Sequence Generation by Cellular Automata”. *Advances in Applied Mathematics* 7 (2): 123–169.

— . 1984. “Universality and Complexity in Cellular Automata”. *Physica D Nonlinear Phenomena* 10:1–35.

Xenakis, I. 1992. *Formalized Music: Thought and Mathematics in Composition*. Stuyvesant, NY: Pendragon Press.

— . 1981a. “Les Chemins de La Composition Musicale”. In *Le Compositeur et L’ordinateur*, ed. by M. Battier, 13–32. Paris: Ircam - Centre Pompidou.

— . 1981b. *Musiques formelles: nouveaux principes formels de composition musicale*. Paris: Stock musique.

Xenakis, I., and O. Messiaen. 1994. *Arts-Sciences: Alloys*. Reprint edition. New York: Pendragon Press.

Ziemke, T. 1999. *Remembering How to Behave: Recurrent Neural Networks for Adaptive Robot Behavior*.

Appendices

Appendix A

Online Documentation of Works

This appendix presents web-accessible documentation of a sample of my work involving adaptive systems which have been discussed in this dissertation, including video documentation

A.1 Absences (2008–2011)

Creator: Sofian Audry

Description of work: *Absences* is an intervention project that involves electronic devices installed in outdoor environments. Taking shape at the frontier of new media and environmental art, it proposes a meditation on solitude and association, interaction and adaptation, natural and artificial, biological and inanimate. Each intervention consists in the creation and installation of autonomous electronic devices in various ecosystems. These artificial agents act within their specific environment. The project is created as an ongoing, residency-based process which was largely site-specific, each context contributing to the conceptual and technical development of the work.

Link to work description: `http://sofianaudry.com/en/works/absences`[1]

Link to project's blog: `http://absences.sofianaudry.com`

Duration of video: 3:25

Link to video: `https://vimeo.com/41576835`

[1] Links to video documentation of individual projects are available at the bottom of this page.

A.2 Vessels (2010–2015)

Creators: Sofian Audry, Stephen Kelly, and Samuel St-Aubin[2]

Description of work: *Vessels* is a robotic installation consisting of large groups of autonomous water vehicles. The robotic agents interact with each other and their environment to form a simple ecosystem. Their collective, emergent behavior resembles the social interactions in a community of living creatures. Observers may empathize with the robots' behaviors, ascribe intentions and motivations to their actions, and/or draw correlations between the group dynamic and unseen characteristics of their milieu.

Link to project's website: `http://vessels.perte-de-signal.org`

Duration of video: 2:34

Link to video: `https://vimeo.com/137104837`

[2]Special collaboration: Adam Kelly.

A.3 N-Polytope (2012)

Creators: Chris Salter, Sofian Audry, Marije Baaldman, Adam Basanta, Elio Bidinost, and Thomas Spier

Description of work: *N-Polytope: Behaviors in Light and Sound After Iannis Xenakis*** is a spectacular light and sound performance-installation combining cutting edge lighting, lasers, sound, sensing and Machine Learning software inspired by composer Iannis Xenakis's radical 1960s- 1970s works named Polytopes (from the Greek 'poly', many and 'topos', space). As large scale, immersive architectural environments that made the indeterminate and chaotic patterns and behaviour of natural phenomena experiential through the temporal dynamics of light and the spatial dynamics of sound, the *Polytopes* still to this day are relatively unknown but were far ahead of their time. *N-Polytope* is based on the attempt to both re-imagine Xenakis' work with probabilistic/stochastic systems with new techniques as well as to explore how these techniques can exemplify our own historical moment of extreme instability.

Link to work description: `http://chrissalter.com/projects/n-polytope-behaviors-in-light-and-sound-aft`

Duration of video: 1:56

Link to video: `https://www.youtube.com/watch?v=hxYJxwfnACU`

A.4 Archipelago (2014)

Creators: Sofian Audry and Samuel St-Aubin

Description of work: *Archipelago* is an interactive audio work involving a group of small electronic modules. The modules emit sounds reminiscent of fictitious birds singing. The songs emitted by the devices evolve during the course of the exhibition using genetic algorithms. Furthermore, they respond to infrared signals, allowing visitors to interact with them using remote controls brought from home or borrowed at the center. The coded messages sent by the remote interrupt and change the songs through crossovers and mutations.

Link to work description: `http://sofianaudry.com/en/works/archipelago`

Duration of video: 1:08

Link to video: `https://vimeo.com/135944076`

A.5 Plasmosis (2013)

Creators: Sofian Audry[3]

Description of work: *Plasmosis* is a site-specific, underwater, artificial life installation. It was installed at the marina of Carleton-sur-Mer (Quebec, Canada) during the Summer 2013. An artificial entity, it interacts in the aquatic environment through multiple sensors, adapting over time to the different natural movements that surround it. The work is thus as a passageway between two worlds: that of air, and that of water. Situated at the point of phase change between liquid and gas, it allows the visitor to exist for a time between these two worlds, to assess their shape, density, temporality and limits. It raises questions about our relationship to the maritime area and its ecosystems.

Link to work description: `http://sofianaudry.com/en/works/plasmosis`

Duration of video: 3:46

Link to video: `https://vimeo.com/86963195`

[3]Electronic and material design: Samuel St-Aubin.

A.6 Accrochages (2008)

Creators: Sofian Audry and Samuel St-Aubin

Description of work: *Accrochages* is a urban electronic intervention project by Montreal-based artists Sofian Audry and Samuel St-Aubin. It stems from their will to bring their art practice out of the walls of a gallery space, on the walls of the city itself. The intent is to build small active and autonomous objects that can, through simple means, give new qualities to the city environment by creating different interactive situations.

Link to work description: `http://sofianaudry.com/en/works/accrochages`

Link to project's blog: `http://accrochages.drone.ws`

Duration of video: 3:39

Link to video: `https://vimeo.com/46397619`

A.7 Vévé (2008)

Creators: Sofian Audry[4]

Description of work: *Vévé* proposes an environment in which the visitor interacts with textual entities through written speech. By taking part in a conversation based on exchange of atomic words, the visitor contributes to the construction of these artificial beings: she teaches them new words, but also new semantic links. Through these poetic dialogues, the entities evolve as their behavior is shaped by interaction with their human counterparts. But there is still place for trials, errors and novelties: in this allegorical space, the artificial creatures often seem to act with their own free will.

Link to work description: `http://sofianaudry.com/en/works/veve`

Link to work: `http://veve.sofianaudry.com`

[4]Audio design: Alexandre Quessy.

A.8 Flag (2007)

Creators: Sofian Audry

Description of work: *Flag* is an interactive dictionary of given names and identities. It allows the visitor to participate by adding his own name and identity to the database. He can also create new names/identities through an evolutionary algorithm. When a visitor wants to add his name to the dictionary, the system asks him to fill out a form about his identity. The form compels the visitor to choose among a limited set of categories and traits, shaping his identity into a socially acceptable, standardized format. Getting back to the evolutionary analogy, if the name acts as the genetic code of the visitor, the traits that form his social definition would be his phenotype. By using artificial recombinations, mutations and crossovers through an evolutionary algorithm, the visitor can then create offspring of his own name or others'. These offsprings' identity traits are recombined and mutated versions of their parents.

Link to work description: `http://sofianaudry.com/en/works/flag`

Duration of video: 3:30

Link to video: `https://www.youtube.com/watch?v=XhIOQm5u7mw`

A.9 CHARACTERS (2005–2006)

Creators: Sofian Audry

Description of work: *CHARACTERS* is an interactive dictionary of given names and identities. It allows the visitor to participate by adding his own name and identity to the database. He can also create new names/identities through an evolutionary algorithm. When a visitor wants to add his name to the dictionary, the system asks him to fill out a form about his identity. The form compels the visitor to choose among a limited set of categories and traits, shaping his identity into a socially acceptable, standardized format. Getting back to the evolutionary analogy, if the name acts as the genetic code of the visitor, the traits that form his social definition would be his phenotype. By using artificial recombinations, mutations and crossovers through an evolutionary algorithm, the visitor can then create offspring of his own name or others'. These offsprings' identity traits are recombined and mutated versions of their parents.

Link to work description: `http://sofianaudry.com/en/works/characters`

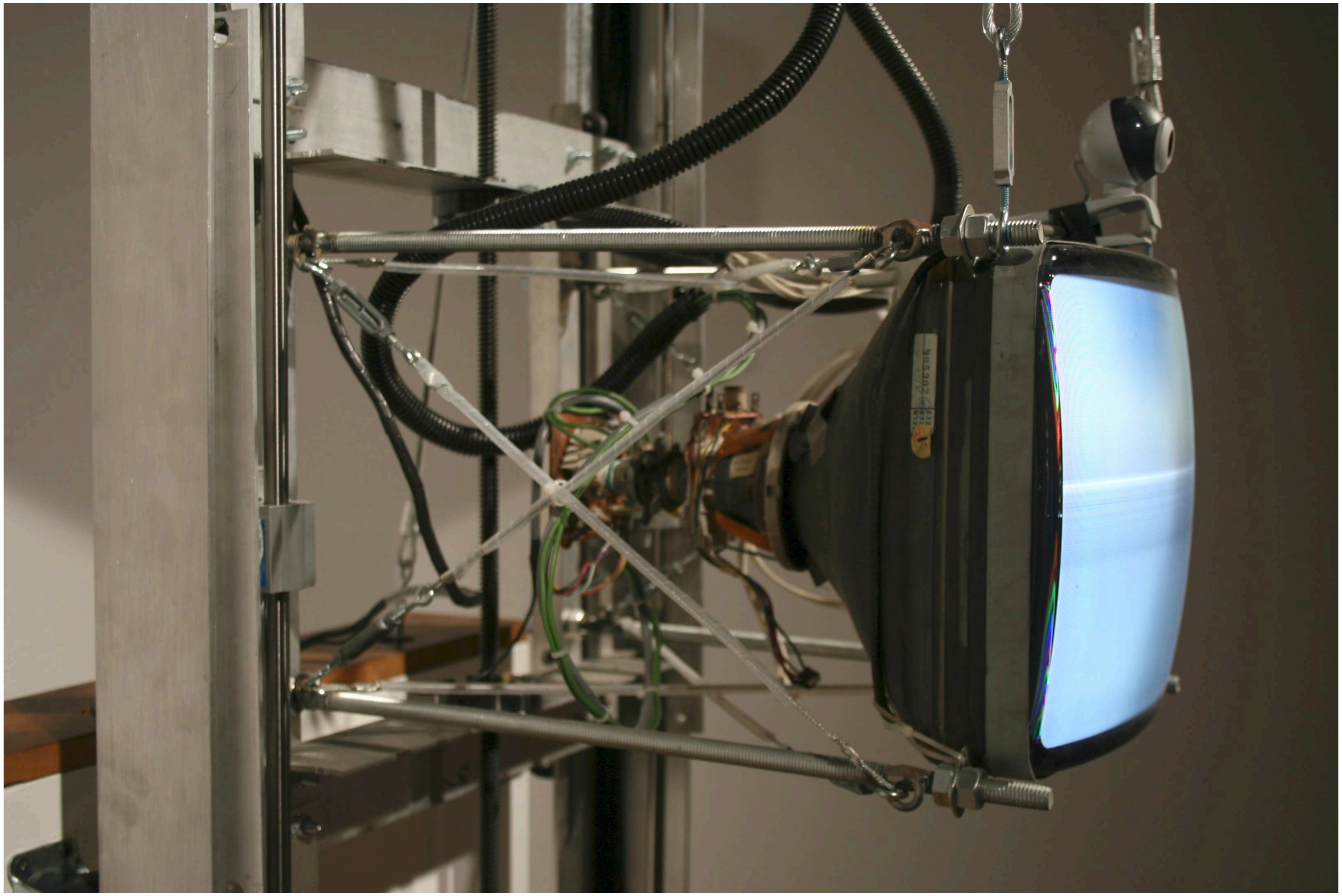
Link to work: `http://characters.tats.name`

Appendix B

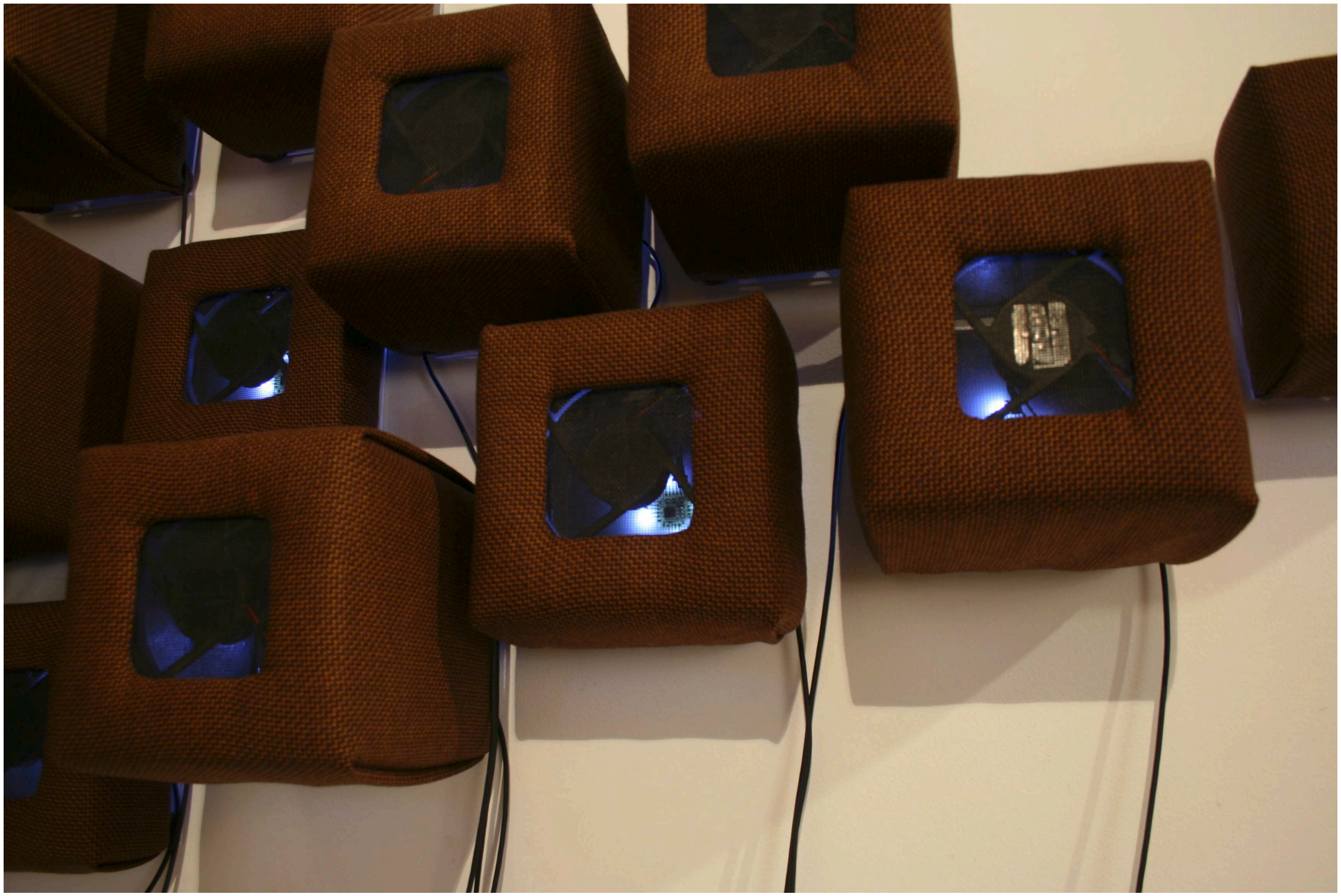
Images of Works

This appendix contains full-page reproductions of works discussed in this research.

Trace L (2007). With Jonathan Villeneuve, in collaboration with Myriam Bessette. Photo by Alexis Bellavance.



Trace V (2007). With Jonathan Villeneuve. Photo by Alexis Bellavance.

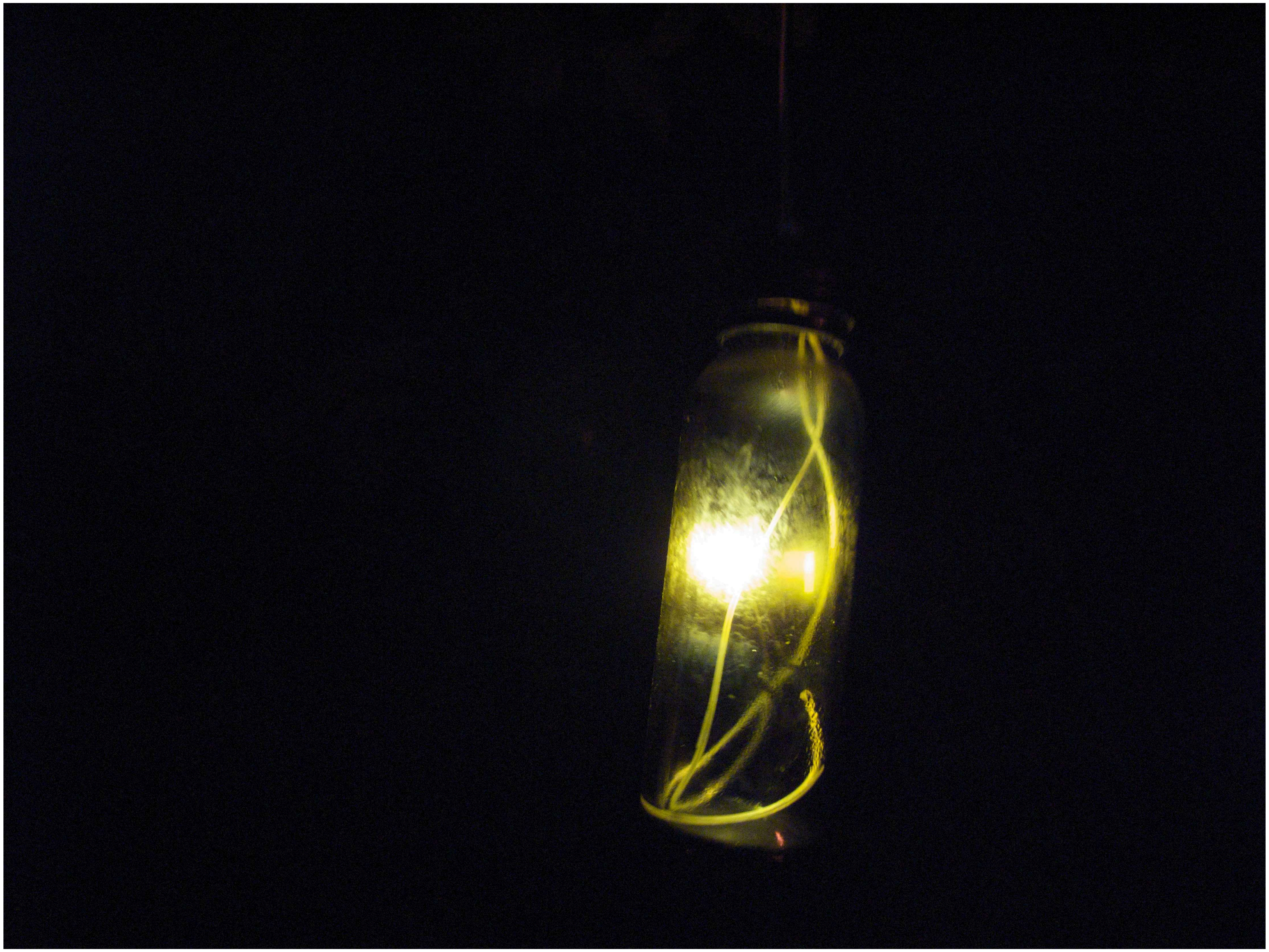


Trace S (2008). With Jonathan Villeneuve. Photo by Alexis Bellavance.

Drift (2007), V2 Institute for the Unstable Media (Rotterdam, Netherlands). Photo by Sofian Audry.

Accrochages (2008). Photo by Alexis Bellavance.

First Absence (2008).

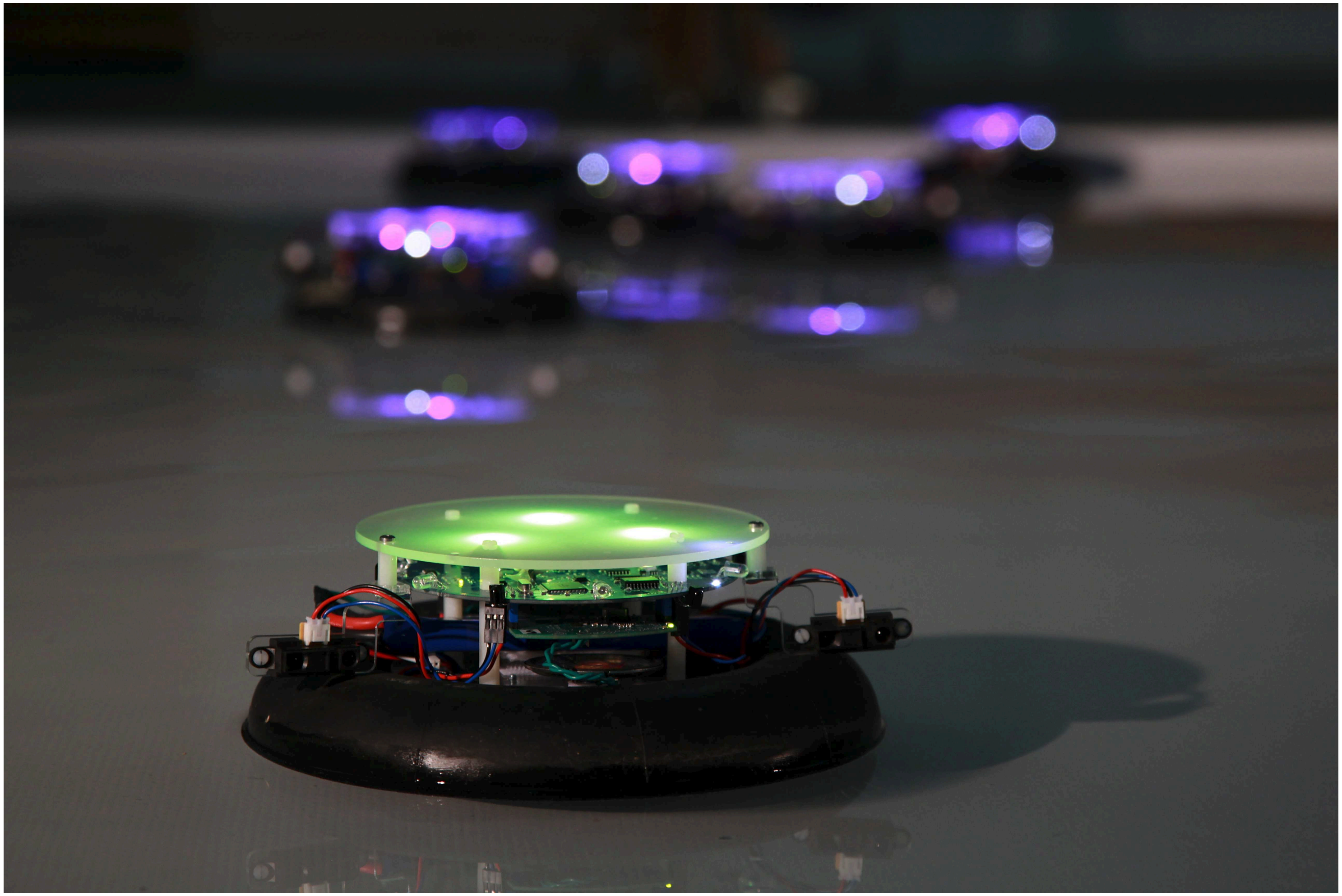


Second Absence (2009).

Third Absence (2010).

Fourth Absence (2009).

Fifth Absence (2011).



Vessels, LABoral (Gijón, Spain) (2013). Photo by Beatriz Orviz.

Vessels, L'Ososphère (Strasbourg, France) (2015). Photo by Philippe Groslier.

Vessels, Nuit Blanche (Montréal, Canada) (2016) Photo by Catherine Aboumrad.

N-Polytope, LABoral (Gijón, Spain) (2012). Photo by Thomas Spier.

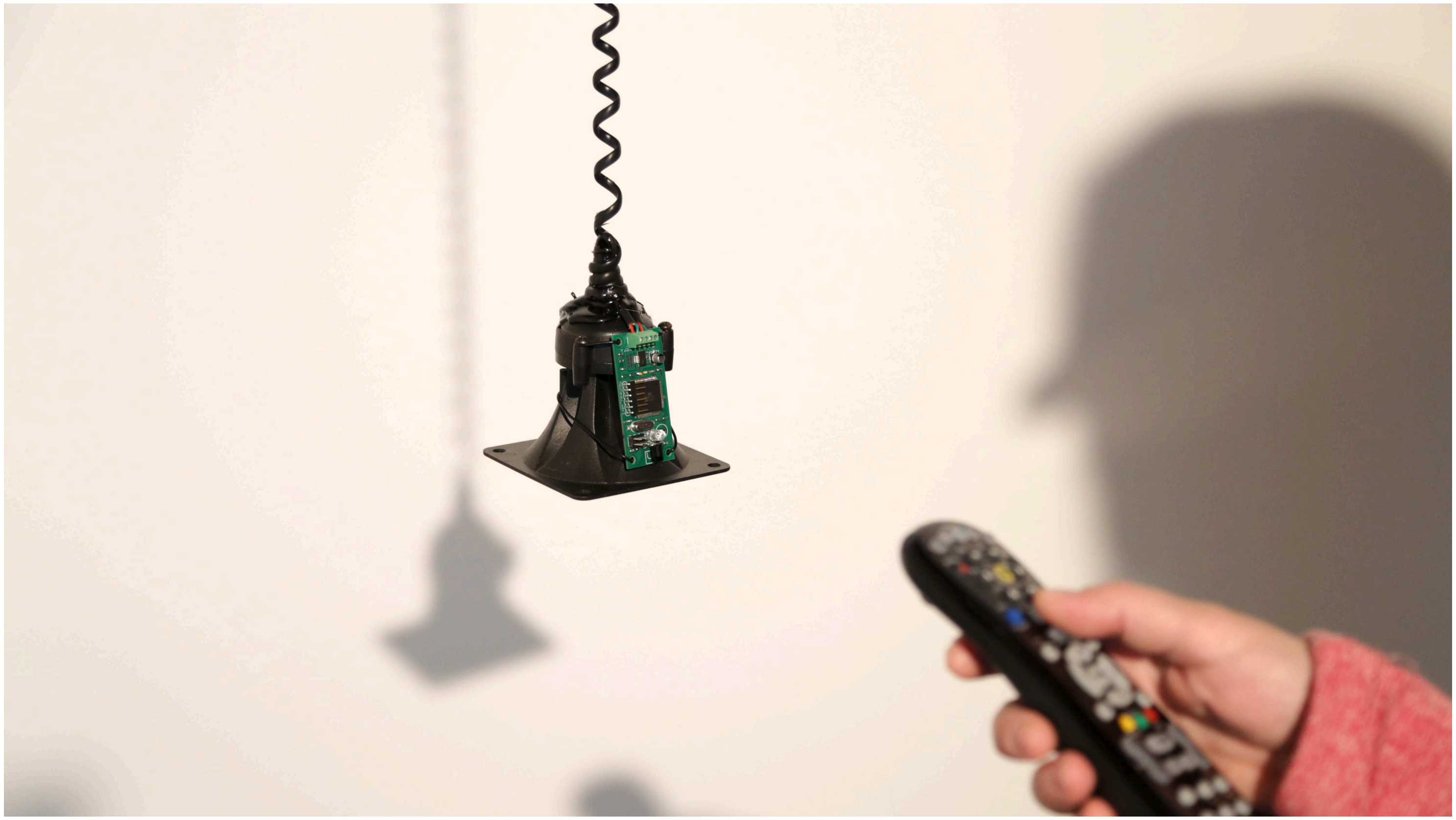
N-Polytope, LABoral (Gijón, Spain) (2012). Photo by Thomas Spier.

N-Polytope, Darling Foundry (Montréal, Canada) (2014). Photo by Thomas Spier.

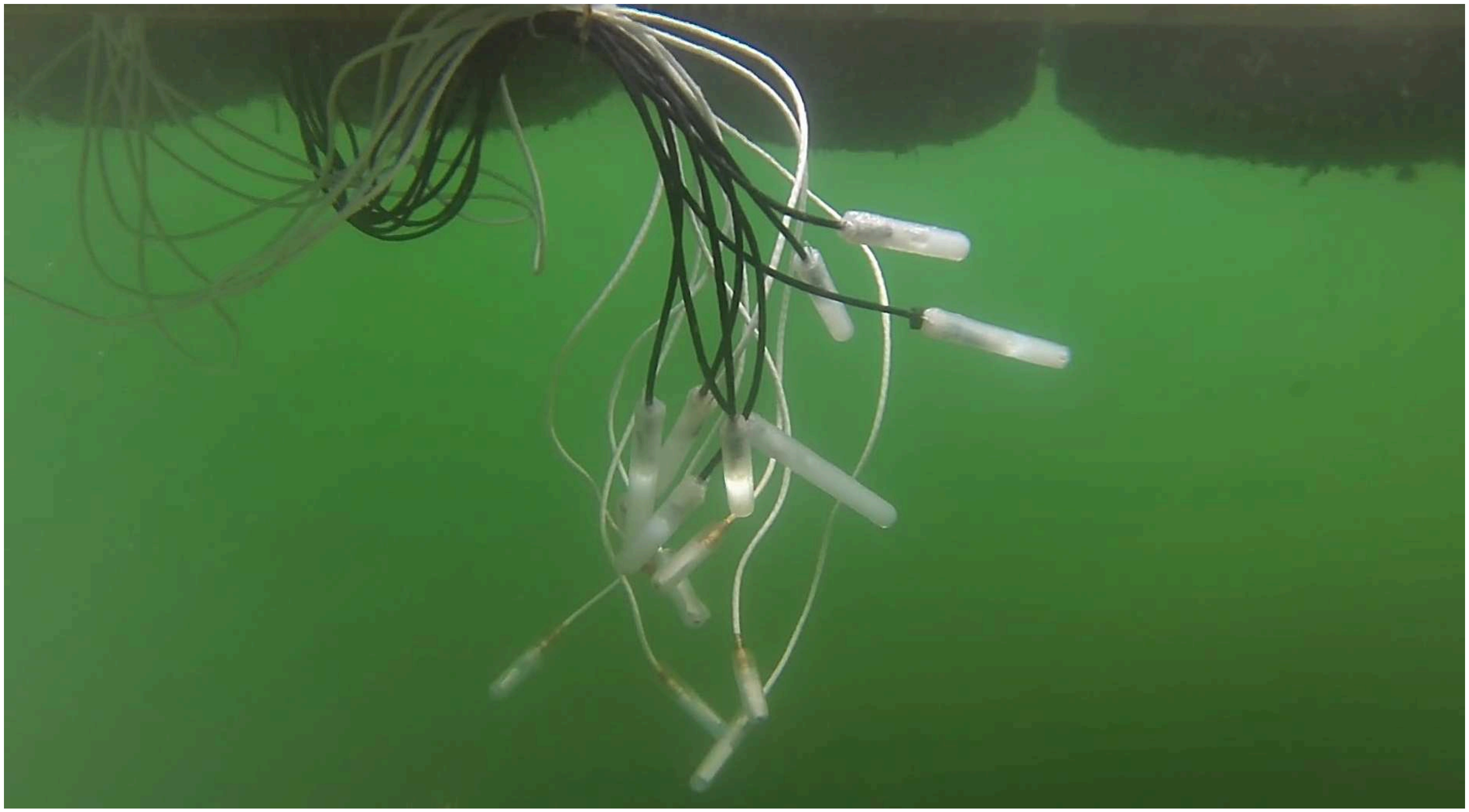
N-Polytope, Darling Foundry (Montréal, Canada) (2014). Photo by Thomas Spier.

N-Polytope, Vitra Design Museum (Weil am Rhein, Germany) (2014). Photo by Thomas Spier.

N-Polytope, Nuit Blanche (Paris, France) (2015). Photo by Thomas Spier.



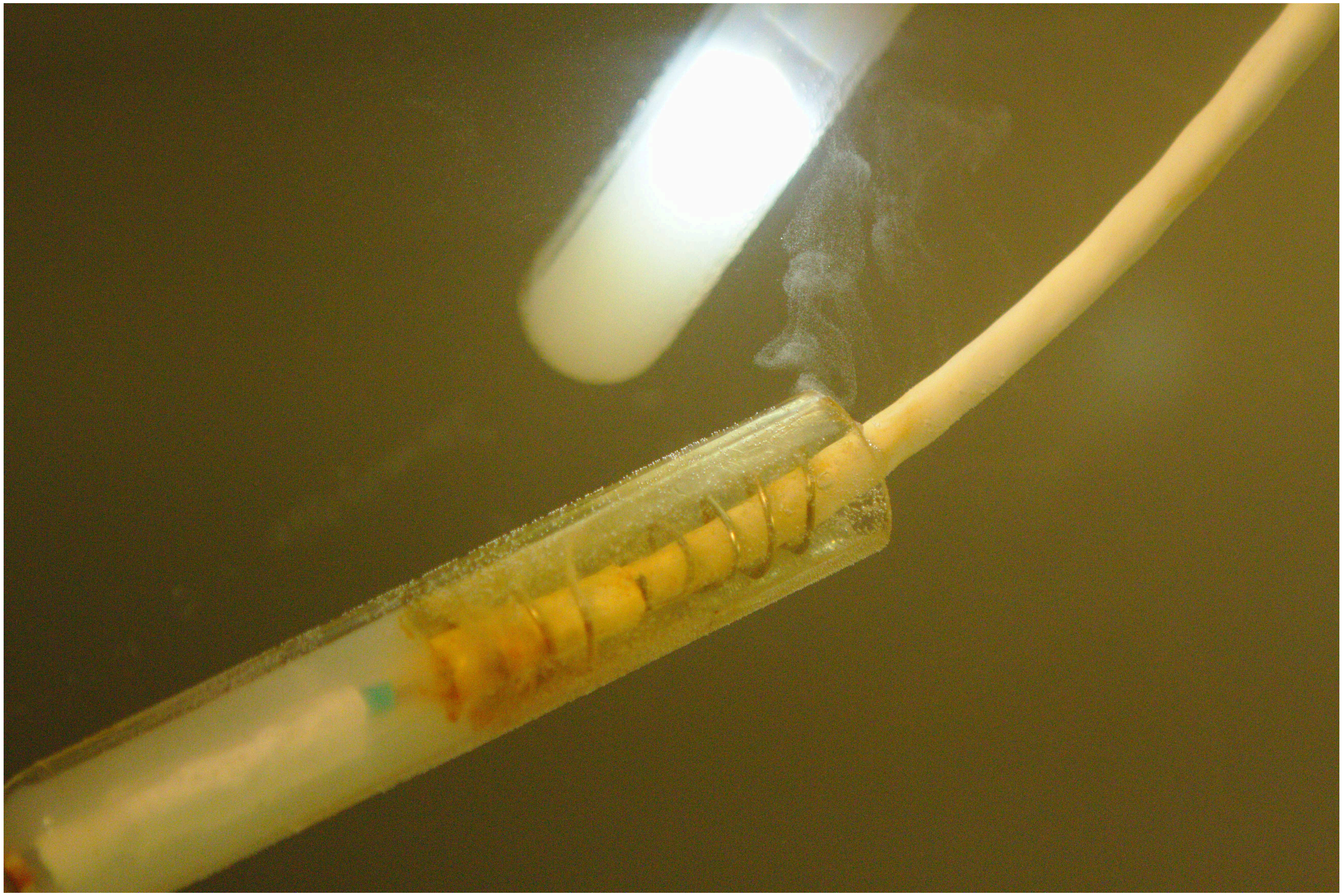
Archipelago, L'Imagier (Gatineau, Canada) (2014).



Plasmosis (2013) (Carleton-sur-Mer, Canada).

Plasmosis (2013) (Carleton-sur-Mer, Canada).

Plasmosis (2013) (Carleton-sur-Mer, Canada).



Plasmosis (2013) (Carleton-sur-Mer, Canada).