Annotated Bibliography

Ben Bogart

May 13, 2011

1 Introduction

Dreaming Machine is a proposed site-specific electronic media artwork that makes explicit use of computational implementations of cognitive models of perception, memory, conceptual development, mental imagery and dreaming. The artwork is to be produced during an art-as-research enquiry (hereafter referred to as "this research") focused on machine origination, machine meaning and the privileged position of scientific knowledge in the construction of our conceptions of humanity. This bibliography reflects a breadth over depth search of related literature. The breadth of the content makes any categorization difficult, and as such each section reflects a theme more than a strict discipline.

Section 2 is placed first as it situates this research in its artistic context and in relation to theories of creativity. *Dreaming Machine* (DM) follows from *Memory Association Machine* (MAM) in that both situated at the intersection of site-specific [7], conceptual [2, 3, 8, 14] and electronic media art [9, 10, 11, 15, 16, 17] practises and are meant to originate visual representations. The research projects differ in that the work on dreaming shifts the context of origination away from the theories of creativity [1, 4, 5, 6, 12, 13] that informed MAM, toward cognitive conceptions of dreaming. That being said, links between cognitive and computational conceptions of dreaming, origination and creativity are expected points of discourse.

In order to bridge the gap between practise and theory, Section 3 presents a compact survey of methodological approaches to art-as-research [18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. This research will be conducted through a reflective practise and formed in triangulation with experts from the various intersecting fields, in particular AI, cognitive psychology and philosophy. The major features of the methodology are: (1) Rigour is applied through continuous reflection that aligns theory and day to day practise. This alignment ensures that the concept (purpose) of the artwork is consistent with the implementation. (2) Validity results from a deep contextualization of the work in artistic, computational and cognitive fields and the faithful application of cognitive models that balance the artistic concept, computational implementation and philosophical/conceptual consistency.

The application of cognitive models requires background knowledge of brain science. Section 4 provides a foundation of brain and neuron anatomy [41, 42], with particular emphasis on dreaming [28, 30, 33, 34, 37, 43] and on memory and visual perception [31, 32, 35, 36, 39, 40, 45, 47]. As the artwork is meant to collect visual data from its context, computer vision techniques are also discussed [29, 38, 44, 46].

Section 5 covers work in infant development and concept generation [48, 50, 51, 52, 53, 55, 56], mental representation [57], and consciousness [49, 54]. Whereas previous work on MAM involved cognitive models of creativity to explore origination, this research looks to how children develop conceptual structures. Additionally, the relation between dreaming and imagination requires knowledge into the nature of mental representation, and perhaps consciousness.

The cognitive models used in this research are to be implemented in a computational system. As such, Section 6 includes various AI approaches covering robotic, connectionist, and cognitive methods, as well as foundational theory [59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70]. This work is partially inspired by Agre [58] in that AI is considered a methodology for revealing conceptual and philosophical issues in underlying theory.

Lastly, Section 7 describes a number of relevant works that are produced in both technical and artistic contexts. Of particular relevance is "Controlled Dream Machine" [75], an Electronic Media Artwork that has "dreams", a number of computational models of dreaming [79, 80, 81], and technically related projects produced in engineering contexts [72, 76]. Generative and metacreation projects are also included [73, 74, 77, 78].

The last paragraph in each annotation is my reflection on the content in relation to this research and its relation to other included works. These discussions are highly instrumental and describe how this research is effected by each citation, and any perceived issues with the work. The phrase "this research" always refers to the *Dreaming Machine* research, not the research discussed in each reference. The phrase "previous work" refers to the research that lead to MAM. Additionally, the use of the word "contemporary" refers to the time the referenced papers were written, except in the case of "contemporary art".

2 Context of Practise

[1] M. A. Boden. Creativity and artificial intelligence. *Artificial Intelligence*, 103:347–356, August 1998.

Boden believes that creativity is a fundamental aspect of human intelligence and that it should be tackled by AI as creative machines could be useful, and give insight into human creativity. Cognition, motivation and emotion are components of human creativity but most computational models focus on cognition. An idea or artifact can be considered creative if it is "...novel, surprising, and valuable (interesting, useful, beautiful...)".

There are two classes of novelty: Psychological Creativity is when an artifact or idea is novel for that individual. If the idea or artifact is also novel for the culture at large then it qualifies as Historical Creativity. Boden distinguishes between three types of creativity: (1) Combinatorial creativity results from the combination of multiple existing ideas or artifacts. (2) Exploratory creativity is the systematic exploration of a conceptual space. (3) Transformational creativity occurs when the conceptual space is transformed, allowing an idea or artifact that would not have been possible in the original conceptual space. Exploratory and transformational creativity are related as the exploration of a conceptual space involves simple "tweaking" of the space, while transformational creativity requires large and significant changes.

Boden cites examples of computer models of creativity: JAPE and Copycat exhibit combinatorial creativity, while EMI, AARON and BACON exhibit exploratory creativity. Systems that exhibit transformational creativity include EURISKO, AM and a number of systems built on Genetic Algorithms.

The core of creativity lies in the evaluation of ideas and artifacts. Even the random generation of ideas and artifacts can be creative if they are evaluated as such. A major problem is that it is difficult to articulate the properties of an idea or artifact that make it novel, useful and valuable. The problem increases when these criteria must be formally encoded in a computational system. Motivational and emotional factors figure significantly in the evaluation of ideas and artifacts "...about which current AI has almost nothing to say." Boden identifies two major bottlenecks in the computational modelling of creativity: (1) Domain expertise is required for the generation of the conceptual space required for exploratory and transformational creativity. (2) Evaluation is necessary and especially difficult for systems that are expected to transform the conceptual space.

Boden's framework was central to the previous work that has lead to this research project. The explicit evaluation of creativity is rejected for one central reason: The value and even novelty of an artifact or idea changes over time and across contexts. What may be considered creative in one culture may not be in another. Additionally ideas and artifacts are open to interpretation. The degree of creativity in an idea or artifact may be different for the agent who created it than the audience who interprets it. For this reason it may not be suitable for the creative system to also evaluate. An idea may be deemed valuable months or years after it originated. The interest in dreaming at the core of this research is an attempt to get away from explicit evaluations and explore the possibility for computational origination where evaluation is implicit and the product continues to be open to interpretation.

[2] B. H. D. Buchloh. Conceptual Art 1962-1969: From the Aesthetic of Administration to the Critique of Institutions. *October*, 55:105–143, 1990.

This paper describes a few of the changing conceptions of art in relation to its form, distribution, method of manufacture, context, use and social value during the 1960s while Conceptual Art was developing. The approaches discussed are not coherent and even contain paradoxical elements. At its core, Conceptual Art is a practise that insists on being outside of the production of formally ordered perceptual objects that constitute traditional art, and do so through a rejection of art history and criticism. In particular the "rigorous elimination of visuality and traditional definitions of representation". The origins of Conceptual Art begin with the use of language in cubist works, in which perceptual structure and linguistic definition are integrated. For Kosuth the aim in Conceptual Art is to abolish the art object entirely in order to replace it with an "art proposition", an artistic idea whose form is simply one manifestation. Early works play in the conflict between the visual structure of an object, and the meaning which is ascribed to it. The readymade is an important precursor where the object is mass produced and becomes an art object due to its recontextualization by the artist: "this is a work of art if I say so". Kosuth's works are intended to be independent of context, and all reference to the world, in short tautologies. This is paradoxical in that the very rejection of the object places Conceptual Art in a artistic tradition, and therefore in a context in which it is to be considered. Objects increasingly break with their traditions as paintings become sculptures and then become instructions and experiences. The 'failure' of Conceptual Art leads to the formally oriented minimalist and pop art of the 70s and 80s, where the hold on the object is "reinstated with renewed vigor."

A central goal in this art-research practise is a rigorous link between the structure of the work (its physical reality) and the concept that informs it. This is highly related to Conceptual Art where the artwork becomes an idea, or "art proposition", which "a machine that dreams" could certainly be considered. That being said, Conceptual Art, as it is discussed in this text, is as much about the relation between the meaning of a work and its form, where in extreme cases meaning could be attributed to any object arbitrarily. According to this conception, the rigorous connection between the object and the concept is in fact a rejection of Conceptual Art, as the object gains significance. The interest in forming a rigorous link between object and concept, through computational methods, is a result of Conceptual Art's redefinition of what art is.

[3] F. Cramer. Concepts, notations, software, art. In *Seminar for Allegmeine und Vergleischende Literaturwissenschaft*, 2002.

This paper proposes a consideration of "Software Art" as a subset of Conceptual Art that involves the aesthetics of code, and an understanding of software as cultural artifact. Computers are often considered "media" as they can pass information between senders and receivers, but this is a reduction as they can also function as senders and receivers. The history of digital and computer-aided art can be seen "...as a history of ignorance against programming and programmers." Software can be Conceptual Art because code is "conceptual notation", it can exist independent of hardware, and it has a history before computers. Cramer describes dadaist systematic poetry as a pre-digital Software Art. Conceptual Art and Software Art share two primary aspects. (1) Both collapse the concept notation and the execution—the artwork is both the code and the process that the code manifests. (2) Both depend on language to encode concepts. Software can be treated as literature, due to its conceptual nature and dependence on language. Cramer accepts La Monte Young's notion of conceptual art: the artwork is in the concept itself and it exists in our minds. He uses the example of the "Composition 1961" which states "Draw a straight line and follow it." to illustrate the overlap between Conceptual and Software Art. The sentence is simple but its execution involves an impossible and infinite task. The same could be said of software that endlessly calculates some value, the description is compact, but the execution may explode to a near infinity of possibilities. Cramer identifies two sides in Software Art: (1) An emphasis on the code itself, through considerations of inherent beauty and elegance. (2) An emphasis on the meaning of the concepts in the software—their cultural relevance. All types of Software Artworks can be situated on the continuum between these poles.

This art-as-research project is centrally concerned with the rigorous connection between the concept and the material of the artwork. The material of the artwork is the process that results from the execution of code. This allows this project to be situated in the context of both Software Art and Conceptual Art. Both the elegance of the process, and its cultural relevance (through the use of scientific models) are core to this work.

[4] L. M. Gabora. *Art, Technology, Consciousness: Mind at large*, chapter Toward a Theory of Creative Inklings. Intellect Press, Oxford England, 2000.

The conceptual space, on which Gabora's "beer can" theory of creativity is dependent, is discussed in detail in this paper. Central to the theory is a conceptual organization of memory where micro-features are located nearby one and other, facilitating associations between similar components. This memory has a two major features. (1) Memory is distributed, but constrained: Memories do not occupy single points in space but are distributed over a small area. If memories did occupy single points in space, if they were non-distributed, the space would be made up of many independent units that are not interconnected. If memories were fully distributed every memory would occupy all memory locations and each memory would loose all individuality. For a network of memories to work, each component must be distributed to a degree. (2) Memory is content addressable: The location of a component resembles the features of that component. A component can then be found simply by searching for the desired features. The combination of distributed (but constrained) and content addressable memory allows components to be retrieved that do not exactly match the features searched for, and similar memories overlap to a degree proportional to their similarity. A result of this distribution is that memories are combined in regions of overlap. Overlapping groups of memory components are concepts. As overlaps form the increasing abstraction leads to higher level concepts that encompass more and more memories. Gabora considers the identification an abstraction as a "creative act".

A stimulus in the world-view causes a "perturbation" where similar memories activate similar memories and so on. The higher the degree of activation the more memories are evoked. The creative "inkling" is a particular pattern of activation that may have never been evoked. During this activation new relationships between concepts may become apparent. Large activations are "streams of thought" that associate many diverse memories and concepts that may only be distantly similar. Gaps between concepts can be bridged by large activations and result in more abstract associations. If the activation is too small, then only a few highly related memories are activated. Gabora correlates a large activation of memory with the de-focused attention and sensitive mental states that are associated with creative thinking.

Gabora describes the "beer can" theory of creativity in greater detail. The qualities of the space, and her notion of concepts, are potentially relevant to the concept generating component of this research project. The model is attractive in that its quite elegant and cognitively oriented. One issue with the model is that concepts are constructed from object features, but it is unclear how abstract concepts, divorced from object features, occur. For example it is unclear what the features of the concept of "justice" are, and what objects it would be associated with. A relation between the implicit knowledge in the memories and an explicit symbolic system is missing. The notion of a concept presented here could be useful in combination with other theories from brain science and developmental psychology.

[5] L. M. Gabora. *Creative Evolutionary Systems*, chapter The Beer Can Theory of Creativity. Morgan Kaufman Publishers Inc., San Francisco, CA, 2002.

This paper proposes a theory of creativity that is cognitively oriented and proposes the core of creative thought is the conceptual organization of memory. The "beer can" theory states that the essence of creativity is not in the individual components but in the structure of the network in which they are situated. Creativity is less about the cans and more about the "plastic thingy that holds them together". The model follows from a notion of cultural evolution where culture is the net result of all human creativity. Evolution is "[a] process wherein a stream of information incrementally adapts to environmental constraints." The "information" is a pattern, or a point in the space of all possible patterns. Patterns adapt through transformation of the space, and are constrained by fitness criteria. Gabora describes an evolving entity as a "sharper mirror image" of the environment, which is composed of many other entities that are undergoing constant evolutionary change. Variation, in cultural evolution, results from the combination, transformation and exploration of cultural patterns (ideas, concepts, attitudes, values etc.). A cultural "genotype" is the network of concepts and ideas that constitute a model of reality, or the worldview, of the individual. The "phenotype" implements and allows the communication of the genotype. Culture is a self-sustained system for the exploration and transformation of cultural patterns. MAV (Meme and Variation) is a computer model in which a society of neural network agents generate ideas through innovation and imitation. In MAV evolution occurs in the absence of imitation but not in the absence of innovation. While biological evolution may be considered a "breadth-first" search, cultural evolution is depth first. Often a particular domain is explored deeply before moving onto another. The creation of culture is highly parallel as individuals explore many domains simultaneously while the culture at large is spread over a large area.

A dominant view is that imitation, though social interaction, is the origin of culture. Gabora considers creativity (innovation) as the origin of culture because imitation requires an already existing pattern. Creativity is required to map out the areas of possible interest, while imitation if useful in the refinement and depth-first exploration of a particular domain. Gabora cites the explosion of brain size, tools and methods in human history as evidence for innovation as the origin of culture. If the ability to imitate was acquired first, then there would be no explosion of culture. The "beer can" theory centres on the world-view of the individual. This world-view is the model of the world that is acquired over the lifetime of the individual. The world-view is the set of memories that are situated a content-addressable space. As more memories are collected they are situated closer together. Eventually memories touch, which results in a general concept that encompasses both memories. The resulting network of concepts is unified and allows the propagation of memory activations anywhere in the network. The process by which one idea leads to another is enabled by the structure of the network, or the "plastic thingy".

Gabora's theory is the central idea that initiated "Memory Association Machine" and lead to the "Dreaming Machine" project. Gabora's theory is oriented toward the generation of ideas rather than their evaluation. The importance of evaluation in metacreation is rejected in this research. Dreams could be considered the ultimate example of unintentional creative generation. Gabora's notion of the world-view as a network of concepts could be useful as the machine is expected to generate concepts from its embodied experience in the world.

[6] A.B. Kaufman, S.A. Kornilov, A.S. Bristol, M. Tan, and E. Grigorenko. *The cambridge handbook of creativity*, chapter The Neurobiological Foundation of Creative Cognition. Cambridge Univ Pr, 2010.

This paper describes current (2010) theories on the neurobiological foundations of creativity. Creativity may be considered the "highest form" of human cognition, or it could be an inherent aspect of cognition. The paper covers two major hypotheses: (1) Facilities for creativity are predominantly in the right hemisphere, and (2) creative cognition is association with the reduction of inhibition of the entire network, which correlates with increased inhibition of the cortex. The definition of creative process is a "...sequence of cognitive operations that give rise to novel insights or ideas." The evaluation of creative products is domain specific and processes and products are dependent on historical and cultural contexts. An idea deemed creative now and for this culture may not hold for another time or another culture. Most research concerning the neurobioloigical mechanisms of creativity divide subjects into creative and noncreative groups, and their differences are studied. Unfortunately there is no subject and domain independent measure of creativity, so experiments cannot be reliably repeated.

The "hemispheric asymmetry" hypothesis states that the right hemisphere is correlated with creativity. This asymmetry is based on work by Sperry on subjects with severed corpus collosums, resulting in disconnection between hemispheres. Subsequent studies have shown that the right hemisphere specializes in "global, parallel and holistic processes", while the left specializes in "sequential and analytic processes". Creative individuals show greater dominance of the right hemisphere. Bogen posits that creativity is not only a result of the right hemisphere alone, but also the left, and the communication between them. For the creative association of ideas a rich collection of ideas is required, and this collection may be best managed by the left hemisphere, which is correlated with "... gathering and storing logical, factual information".

According to the disinhibition hypothesis, cortical inhibition is correlated with the ability to shed the schematic constraints that impede creative associations. Creative cognition involves a low activation of the cortex. This cortical arousal changes over time, where creative individuals show low activation during creative tasks and normal activation during non-creative tasks. Studies have shown that creative individuals share a tendency to exhibit a number of features: they may be overtly reactive, show higher GSR values in response to auditory tones, show poor reaction times when distracted, show higher basal activation (and higher variability of activation) of the cortex and are worse at controlling their own EEG patterns. Creativity may be emergent and result from "unintentional inspiration". Paradoxically, creative people can tend towards hermetic life, and yet also seek novelty. This may be explained if disinhibition is considered the originator of creativity, while hemispheric asymmetry allows a creative idea to be executed and evaluated. Studies have shown genetic and behavioural links between creative subjects and those with various developmental illnesses, such as ADHD, schizophrenia and bipolar disorder.

The conception of the "split brain" above appears not to correlate with Zhang's characterization of "representational" and "symbolic" hemispheres. Possible connections between representation and analytic thought, and between spacial representations and originative thought should be explored. The low activation of the cortex corresponds to Gabora's conception of a large activation in the "conceptual space" and high cortical activation corresponds to small (analytic) activation in the conceptual space. It is not clear how Gabora's theory could relate to hemispherical asymmetry. The distinction between right dominate origination and left dominate evaluation is relevant to the relation between dreams and creativity and begs the question—do dreams show hemispherical dominance?

[7] M. Kwon. *One Place After Another: Site-Specific Art and Locational Identity*, chapter Genealogy of Site-Specificity. MIT Press, 2004.

Site-specific art emerged in the late 60s and early 70s with an emphasis on the physicality of the artwork. The work was considered a unique combination of physical elements in a particular place in physical space. This emphasis partly results from a rejection of the gallery context, considered a modernist and "pure idealist space". This emphasis on physicality moves the intellectual reception, the instant "epiphany of the disembodied eye", to a physical engagement with the here and now in real space and real time. This relocation of meaning from the object to the relation between object and context rejects the commodity status of art, as it is not easily sold or collected. These artworks are inextricably linked with their sites, and according to Robert Barry and Richard Serra, "to remove the work is to destroy the work." Eventually this minimalist conception of the site shifts: the context of a work becomes cultural, rather than physical. According to Burren, artwork that does not reflect on the invisible institutional framework of the gallery site engages in an "illusion of self-sufficiency"—the artwork requires no relations outside the aesthetic experience. According to this view, the purpose of a work located in an institutional site (in the cultural sense) is the recoding and decoding of conventions for the purpose of exposing hidden operations. In the 1980s work moves beyond this institutional sense of site and expands to include social, economic and political contexts. Contemporary site-specific art often rejects visual aesthetic pleasure and aims for a more intense engagement with the public, through happenings and other public interventions. The site becomes not a physical location, but a intersection of "texts"—various contexts of meaning. The transformation of site-specific art slides between notions of physical location (grounded, fixed and actual) to discursive (ungrounded, fluid and virtual), reflected in the shifts between phenomenological, social/institutional and discursive modes.

The precursor to this work is explicitly considered site-specific art. In that case, the notion of site-specificity is a physical one, meaning the location in which the artwork is installed. The link to site results from the causal processes encoded in software. For this research, the physical consideration of site is still in play, but the work also is site-specific in its cultural space. Partly inspired by Agre, the cognitive, computational and psychological placement of the project is the intellectual site of production. The project explicitly requires that these areas are balanced in order to ensure validity. The purpose of placing ideas in these domains in an artistic context is to reflect on them in relation to each other. The artwork is not simply the object/process but also the intellectual development of the project, through triangulation and discussion with specialists, and documented through logs and the dissertation. The intended physical installation is meant as a site of questioning and considering what it means for a machine to dream, for AI specialists, cognitive psychologists, developmentalists, artists and the public at large.

[8] S. LeWitt. Sentences on conceptual art. *Alberro and Stimson*, 1999.

This work is a set of 35 sentences that define Conceptual Art for Sol Lewitt. Rather than paraphrasing or quoting, points relevant to this research project are extracted. In Conceptual Art, irrationality and mysticism are favoured over rationality and logic. This research project overlaps with AI, but rejects rationality in order to explore computational models of phenomenon that transcend it, in particular dreaming. For Lewitt, logic should be used to explore an irrational idea. If dreaming is irrational then the use of computational and cognitive models to manifest them is consistent with Conceptual Art. The process has a life beyond the artist's will, which "may only be ego". The emphasis on process over object is central to this art-as-research, and the underemphasis of will relates to meta-creation and the desire for a machine to surprise the artist. In Conceptual Art, an idea itself may be a work of art, without ever being manifest. Processes and matter are equal opportunities to manifest conceptual art ideas. The purpose behind this research is an artistic proposition, the idea of a machine that dreams, although the machine will actually be constructed. Any idea that is concerned with art and fall within artistic conventions may be considered an artwork. The computational system produced during this research will be an artwork, but its material is simply software, a set of processes. One role of art is the alteration of the conventions of art. This art-as-research project is as much a reconsideration of what art is as it is a reconsideration of research, computation, intelligence and scientific knowledge. Even if an artwork can just be an idea, that idea cannot be reflected upon until it is materialized in some way—for example stated in language. Lewitt believes that the interpretation of the artist is no more valid than the interpretation of the viewer—The artwork is open-ended.

[9] J. McCormack. New challenges for evolutionary music and art. *SIGEVOlution*, 1:5–11, 2006.

This paper describes important research directions to follow over the next fifteen years in evolutionary music and art (EMA), which is considered both an artistic and scientific form of enquiry. McCormack provides two broad categories of EMA based on goals, which he considers more significant than the disciplines of the practitioners: (1) Art-making/understanding systems are intended to make art or music that is evaluated and appreciated by a human audience. Most existing EMA projects lie in this category, typically make use of evolutionary processes and depend on domain specific knowledge. (2) Artificial creative systems focus on the concept of creativity in general. Creative behaviour that is not strictly human, art-as-it-could-be, is explored.

The central problem in art-making/understanding systems is the search for an interesting phenotype in a very large search space. In this category production has two primary considerations: (1) The design of the generative system and its parametrization, and (2) the evaluation of the phenotypes produced by the system. Often a particular system will only construct images of a particular aesthetic class. In order to break out of this restriction a "system capable of introducing novelty within itself" is required. Ideally a system would implement a bottom-up and non-teleological process of selection, self-assembly and self-organization. The genotype and its interpretation should be part of a unified system.

In aesthetic selection a human acts as the fitness function and chooses which phenotypes are successful. This methodology does not tell us much about creativity in general and is problematic because: (1) The population size is limited by the ability of the artist/user to select one phenotype from a large set. (2) Selection is slower than the computation of phenotypes. (3) Small changes in the genotype can lead to large changes in the phenotype. (4) The size and complexity of genotypes are limited. An alternative is to allow the system to do phenotype selection. This is a complex and open problem because it is not clear how to measure aesthetics, as properties are interpreted in a context sensitive manner. The role of the environment is often overlooked in EMA systems, where we should expect the complexity of the environment to be reflected in the complexity and detail of the phenotypes. Human creative endeavour happens in a social structure, but few EMA systems incorporate cognitive and social dimensions in the system.

EMA has the potential to provided an "extended interface" that may rival the expressivity and flexibility of traditional tools such as pencils and cellos. The writing of software is a primarily conceptual exercise and lacks "immediacy, physical tactility and environmental interaction." For an extended interface to be successful the artist/player must feel immediacy with the system and the system must be used by many, mastered by few, and becomes the subject of study in art and music institutions.

Although this research project has not yet taken a position on the use of evolutionary algorithms, the classification of EMA goals can be applied beyond EMA. There is nothing specific to evolutionary methods in "art-making/understanding" and "artificial creative" systems and both could be implemented using any number of generative methods. This research lies somewhere in between. The "dreams" of the machine are not considered art in themselves, although they have been exhibited as standalone documentation, and are expected to be interpreted aesthetically by the audience. Simultaneously the artwork is in the existence of a "machine that dreams" and the specific process it implements. The process is meant to be considered conceptually by the audience and leans more towards the artificial creative category because it leads to dreams inspired not only by human experience. The restricted embodiment of the system makes its "experience" different than that of a human's. The system should produce dreams in a free manner, unconstrained by explicit evaluation. Artificial creative systems are meant to explore creativity in a cultural and species independent way, which conflicts with Boden's conception of creativity that centrally requires evaluation, which is cultural and species specific.

[10] L. Poissant. *Media Art Histories*, chapter The Passage from Material to Interface. MIT Press, 2007.

Possiant describes the shift in contemporary art from material to interface through an analysis of the historical precedent and function of interactive arts. Much of what is taken in art history is fractured by interactive art. This fracturing of history is not new, but has often coincided with technological changes. Art is constituted by three "poles": (1) In traditional art the artist is committed to representing their vision of the world by expressing their emotions, a picture of reality, spiritual and political values, etc.. (2) Changes in artistic material reveal changes in the roles of the artist—visionary, creator, "denunciator", consciousness-raiser, absorber of an era's sensibility, etc.. (3) New materials point to the reconfiguration of the relation between artist and viewer: The viewer is called upon to intervene directly in the process of creation. The materials and tools we use effect the enquiry we engage in—"[t]here is no innocent material". Transparent materials and tools "metabolize" knowledge, effecting how we see, sense and think.

The intellectual climate in which the shift from material to interface arises is comprised of a number of influences. Wittgenstein, Austin and Armengeaud conceive of language as action. For Eco interpretation is also an action as the viewer participates in meaning-making. The growing importance of the role of the reader changes the art's base from a "petrified solidity" to a "nervous mobility or relations and connections". This marginalizes the artist's privileged position as meaning-maker and emphasizes a relation to the viewer. Artistic precedent comes in the form of attempts to integrate audiences into the "show"—for example, by using them as a projection surface as early as 1909. In the 1960s artists and the public engage directly in happenings. In early interactive video art, produced by Paik, Levine, Nauman and Cincera, the viewer directly manipulates images.

Interactive art is composed of six classes of "conductors" that allow the viewer to engage in the artwork: (1) Sensors read the state of the world, or itself. (2) Recorders store aspects of reality in variously time-stable methods. (3) Actuators cause physical changes in the world. (4) Transmitters send information over distance. (5) Diffusers project information to be sensed into the world, eg. displays, speakers, etc.. (6) Integrators are artificial creatures that are attempts to reproduce the living. These conductors can be used to accomplish the five functions of interactive art: (1) Extend: senses are transformed for reach and/or accuracy to be lengthened. (2) Reveal: interfaces reveal that which we can not normally see or "objectify". (3) Rehabilitate: the forgotten or neglected senses are restored. (4) Synesthetic Integration: Senses are crossed such that colour becomes sound, or touch becomes colour. (5) Filtering: Methods of reducing the amount of information allow inter-actors to orient themselves. Interactive art allows the probing of space to explore new territory. Traditional art was meant to be thought about and not felt, or to be felt in an intellectual fashion. Interactive, "communicational", arts privilege the multiplication of channels of reception.

The artwork produced in this research is not intended to be overtly interactive. The viewer is not meant to form an immediate engagement with the work. As these works are meant for long term exhibition, in public space, they are expected to form a relationship that develops over time. Viewers are meant to live with the works for months and years. The relationship between the work and its environment mirrors the link between an interactive artwork and its viewer/inter-actor. The same technical methods are used but they are not focused on the viewer. The artwork and the research are not isolated components but form a larger unified entity that is meant to be shared. The integration of theories from neurology, cognitive science, computation and relates this work to a "communicational" practise.

[11] D. Rokeby. *Transforming mirrors: subjectivity and control in interactive media*, chapter 7, pages 133–158. SUNY Press, 1995.

David Rokeby discusses the historical precedent of interactive arts, four models of interaction, and their relationship to a media's ability to support subjectivity and control. Interaction is about a technology's ability to reflect human behaviour back at ourselves. While engineers attempt to make media transparent, artists use the transformations inherent in media to "explore the meaning of the interface itself". Rokeby describes two notions of interactivity in art: (1) The act of interpretation is a type of interaction as the viewer participates in meaning making. (2) Part of this interpretive power is displaced from the viewer directly into the medium. The former concerns static art, while the latter is the case in overtly interactive art. If a creative process is the iterative removal of possibilities, then a static artwork results from a reduction of options until there are none. In interactive art, the process is stopped part way such that the remaining choices are left to the viewer to determine. This is similar to Cagian chance operations, where the final choices are left to randomness. Perhaps true interaction requires that the artwork "...include[s] some sort of adaptive mechanism, an apparatus for accumulating and interpreting its experience."

Navigable structures are spatial compositions created by the artist that are explored by the viewer. These compositions can take on any number of geometries, from three dimensional to tree-like. The more constraints to this navigation the more the "sense of personal impact". Systems with few constraints are often unsatisfying. In a static artwork, the illusion of freedom is restricted while the possibilities for interpretation increase. In the case of interactive artwork the opposite is true, a restriction of interpretation includes an increase in the illusion of freedom.

Media artists may create artworks that are the invention of media. This is the case in interactive systems that allow the viewer to express their own creativity, for example in Myron Krueger's "Videoplace". The viewer is free to play creatively because the constraints in the system protect the viewer from the fear of their incompetence. The engineer's interest in transparent media contrasts with artists interest in expressing ideas through its opacity. The relations possible through the media are more significant than visual and aural aesthetics.

Transforming mirrors reflect the viewer's behaviour back at themselves. Sometimes a transformed image of the viewer herself is used and recognized. Rokeby references the myth of Echo and Narcissus describing the role of the artist: "the interactive artist transforms what is given by the interactor into an expression of something other." These transformations become meaningful through the tension between the viewer's sense of self and the responses of the system. The illusion of power one gains through an interactive system is an illusion because it is highly constrained and lacks responsibility. In this kind of work the desire for the system to surprise its creator is common.

Automata are not interactive in the same sense as these other models. These systems explore their world, of which the viewer is "only one aspect". The ultimate goal in this type of work "... is the self-replicating, self-sustaining machine—artificial life." These works do not reflect the individual interactor but reflect conceptions of human behaviour. For artists such as Norman White, these artworks are a method for self-understanding. Often these systems aim to exhibit emergent properties that allow the "transcendence of the closed determinism implied by the technology and the artists' own limitations."

The interactive artist does not create finished works but creates systems of relationships. This gives the artist expressive powers: "The works are somewhat akin to portraits, reflecting back aspects of the interactors, transformed so as to express the artist's point." This act of expression moves to a higher level of abstraction, manifested in the media rather than the content. Rokeby warns about the apparent transparency of interfaces: "If we allow ourselves to lose consciousness of the influence of the interface, we lose our ability to question the terms of the contract; the contract will be effectively invisible. If we accept the transformed images reflected back through the mediating technologies as images of ourselves, we surrender the ability to control who we are."

Rokeby is a primary influence on the artistic practise in which this research is situated. The overlaps in concern and interests are nearly enumerable. Central concepts are the conception of the automata, the desire for surprise from a constructed system, the consideration of interactive art as a type of portraiture. The "Context Machines", a series of works of which "Dreaming Machines" are situated, expand Rokeby's conception of automata and mirrors to the landscape, a transformed representation of a particular place in space and time. The explicit use of cognitive models in this research support Rokeby's conception of automata as mirroring our conceptions of human behaviour. It is not just a reflection of human behaviour, but when considered in light of A-life, a mirroring of conceptions of life in general.

[12] R. Saunders and J. S. Gero. How to study artificial creativity. *Proceedings of the 4th conference on Creativity & cognition*, pages 80–87, 2002.

This paper provides a methodology, called "Artificial Creativity", for studying creativity through the use of a computational model of Csikszentmihalyi's systems view of creativity. It is meant to explore the nature of creativity, both as it is and as it could be. The methodology allows creativity to be seen through lenses of cognitive science, sociology and history.

Csikszentmihalyi's view of creativity is based on a society of agents that interact through a social structure. The conception is composed of three major modules: (1) the individual is an independent creative agent, (2) the field is the social context of creation and (3) the domain is the long-term cultural storehouse of the products of creativity. In a typical cycle, knowledge in the domain is transformed by individual production. If the results are considered worthy by the field, then it is included in the domain, ready to be transformed by another individual. Most study of creativity focuses on one of these modules, and not on the whole system.

The Artificial Creativity methodology is composed of the following features: (1) The model contains agents situated in a cultural environment. (2) No single agent can direct the behaviour of other agents. (3) No rules dictate global behaviour. (4) Agents can interact through the exchange of artifacts and evaluations. (5) Agents access cultural symbols from the environment, and (6) they evaluate the creativity of other agents. Features are inspired by A-life methodology, except the ecological environment is replaced by a cultural one and agents evaluate each others productions. The criteria for judgement is individual and social judgements emerge from their interactions. The methodology is applied in the "Digital Clockwork Muse", a simulation of a society of "curious design agents", and "explore[s] the role that an individual's search for novelty plays in socially situated creative systems."

A "curious design agent" is designed to both construct genetic artworks, and to evaluate those constructed by itself and its peers. Each agent has a small SOM that is used to determine the novelty of artworks. Novelty is determined from a distance between the most similar of the SOM prototypes and the evaluated artwork. This measure of novelty is transformed into a measure of "interestingness" through the use of a Wundt curve, where artworks showing low and extreme novelty are deemed uninteresting while artworks with just the right amount of novelty are deemed interesting. The "novelty preference" corresponds to the amount of novelty at the point of maximum interestingness and varies between individual agents.

In experiments, agents with a low preference for novelty innovate slower than those with high novelty preference. Creativity is attributed most to agents in a society whose novelty preference is near the average. Agents with high and low novelty preferences are attributed with little creativity. In cases where agents are divided into two groups, by their preference for novelty, they tend to attribute creativity within, and not between, groups. Each of these groups tend to construct artworks that are perceptually distinct, where each explores an independent subspace. In cases where creativity is attributed by one group to another, the relation tends to be asymmetrical—one group sees the other as creative, but not vice versa. The perceptual separation between groups is reflected in a measure of the fractal dimension of the artworks and is easily distinguished between groups. The degree of complexity, as measured by the fractal dimension, correlates linearly with the novelty preference of the group. Groups with high novelty preference tend to create artworks with a high degree of complexity.

This research project is focused on an individual dreamer, and not on a social structure of dreamers. The methodology proposed above is relevant as it does not depend on any absolute measures of creativity or novelty but only relative measures specific to each individual. The success and failure of creative productions is an emergent result of the cultural interactions between individuals. This approach to evaluation in meta-creation is highly attractive as it does not involve the apriori determination of evaluation criteria.

[13] J. Schmidhuber. Simple algorithmic theory of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. *Journal of the Society of Instrument and Control Engineers*, 48(1), 2009.

Schmidhuber defines notions of subjective beauty, interestingness, creativity, etc. in the context of our ability to reduce the structure of sense patterns into compact ("compressed") representations. The process of scientific development is to break the world into manageable pieces and determine laws that explain the structure of those pieces. This reduction of phenomena to laws is a process of compression where the laws form a compact representation that can stand in for the phenomena. Curiosity drives artists and scientists to stimulus that contains unknown, but knowable structure.

The algorithmic framework is composed of four major principals: (1) All input patterns should be stored in the system. It is argued that this is plausible for the human brain. (2) Improve "subjective compressibility" by using any regularity (structure) in the input. (3) Let curiosity reward reflect the compression progress, measured by the number of bits needed to encode historical (previously collected) data. (4) Maximize the intrinsic curiosity reward through reinforcement learning. This allows the system to focus on "finding or creating new, previously unknown but learnable regularities."

Schmidhuber's conception leads to a number of consequences for high level concepts: (1) Compressed representations are "symbols". (2) Consciousness is defined as an internal symbol for the agent to represent itself. When this representation is used then "the agent could be called 'self-aware' or 'conscious.'" (3) Subjective beauty is a function of a particular observation, at a particular time for a particular agent, and is proportional to the size of the compressed representation. (4) "Interestingness" is not the same as, but is the first derivative of, beauty. A novel stimulus may be interesting, but not beautiful because its structure is not yet learned. As the compressor increasingly learns this structure, the beauty of the stimulus increases. Eventually, the compressed representation shrinks, which reflects a high degree of beauty, but as the structure is known it is no longer interesting. (5) If the compressed representation of the stimulus does not shrink, then it is no longer surprising because its structure is not learnable by the system. (6) In the absence of external reward the system maximizes interestingness in order to seek new and learnable stimulus. Schmidhuber goes on to describe how these concepts relate to Art, Music, Science, Comedy etc., which are all motivated by this curiosity driven exploration of compressible stimulus.

The ideas proposed in this paper are highly relevant to this research due to their emphasis on self-motivated behaviour, which is a core philosophical requirement in this research. Additionally the conception of a compressor as representation appears to relate to concept development. The essential structure of a stimulus is extracted through rules learned through the compression of previous stimulus. Each new stimulus is then considered in light of previous stimulus. The combination of curiosity, beauty and interestingness can be used to motivate an agent to explore and learn from the world in an unsupervised fashion. The argument that the human brain could store the entire stream of lived sensory experience is not convincing, and the limitations of storing every stimulus is problematic. If we consider these compressed representations as concepts, it is not clear how this system deals with the relations between concepts, are they hierarchically organized? Can an associative process flow through them? Are the rules that map stimulus to compressed representation implicit or explicit? Nevertheless to approach could be applied to this research.

[14] C. Tamblyn. Computer art as conceptual art. *Art Journal*, 49(3):253–256, 1990.

This paper provides a survey of early "computer art" produced during the 1980s. Works are selected when they do not use computers to emulate other media, but are artworks that could only exist on computers. The core features of conceptual art are: (1) the rejection of the object, (2) that art's purpose is to question art, and (3) where audience participation breaks down the relation between artist and viewer. Computer art directly follows from these concerns. Works discussed can be grouped into two broad categories: (1) Narrative projects, such as "Deep Contact" by Lynne Hershman and "Margo" by Sara Roberts and (2) Interactive Games, such as "Who Says" by John Manning, "Composite Machine" by Nancy Burson, "SYMulation" by Ed Tannenbaum and "Warp It Out" by Jane Veeder. The narrative projects are hypercard based interactive video disk installations that make use of cinematic conventions in an interactive context. The interactive games allow the viewer to use their own image as the basis of image-processing operations. For Tamblyn, computer art is the "ultimate conclusion" of the Duchampian agenda to explore the "ephemeral and contingent aspects of meaning".

This art-as-research practise is situated in the continuum of conceptual art, for the same reasons as summarized above. Centrally the practise questions the role and purpose of art, and what constitutes an artwork, as well as the very nature of meaning itself. This paper enriches the contextualization of electronic media art practise, not as "new" but as continuation of existing practises.

[15] M. Whitelaw. *Metacreation: art and artificial life*, pages 1–22. The MIT Press, 2004.

Whitelaw provides an introduction to the A-life art field and a typology of works. A-life science is the creation and study of artificial systems that mimic or manifest properties of living systems. A-life art uses techniques from A-life science but situates the work at the intersection of contemporary art practise and new media art.

A-life science subscribes to materialism where the living things are considered wholly physical biological machines. The premise is that life results from a dynamic structure of matter that manifests itself primarily through its behaviour, or physical actions. This consideration of life makes the simulation of life a valid enquiry, as life is composed of abstract material processes. The importance of material and physical processes emphasizes the "bottom up"

approach where global complexity arises from the simple interactions of many components—emergence. This contrasts with artificial intelligence where a "top-down" approach is common, and is problematic for A-life systems.

Whitelaw describes four key A-life techniques that correspond to his typology. Genetic Algorithms are simulations of biological genetics. Reproduction and mutational allow a simulated evolutionary process that is constrained by fitness criteria. This method can be used to solve functional problems and corresponds to Whitelaw's "Breeders" type. Agent-Based Systems are composed of many components that interact in an artificial world. Agents range from simple (exhibiting behaviours such as eating and breeding) to complex (exhibiting behaviours such as communicating and cooperating). These systems allow the possibility of predator/grey relationships and can exhibit the same kind of bottom-up complexity as "Breeders". These agent-based works correspond to the "Cybernatures" type. Bottom-up Robotics make use of simple parallel mechanisms that result in complex higher level behaviour, as in Brook's "subsumption architecture". These robots fit into the "Hardware" type. Cellular Automata are purely formal explorations of emergence. They involve arrays a cells whose rules determine their future states based on the current state of their neighbours. Works that use similar techniques correspond to the "Abstract Machines" type.

A-life science methods have been explored by artists since the introduction of the field in 1987. Prominent early adopters include William Latham and Carl Sims, who followed biological and computational sciences. They showed early A-life artworks at prominent events and festivals in the 1990s. A-life art is often correlated with interactive art as it provides an elegant method of actively engaging the audience. A-life also includes non-interactive aesthetic artifacts that arise through the artists engagement with A-life processes. Whitelaw describes a number of artistic precursors to A-life art. Goethe focused on the dynamic whole of a natural object over its constituent parts and advocated that art must be true to nature through the study and imitation of natural processes. Klee and Malevich's abstractions where inspired by the invisible forces of nature. Artists were actively engaged in Cybernetics, characterized by an interest in abstract causal dynamics, input/output systems and feedback loops, after world world II. According to Burnham, this marks a shift from the art object to the process in sculpture. A-life art is a recent extension of a rich tradition of practise concerned with the appearance and functional structure of nature through imagined and applied technological means. These practises point to a teleological direction of art, toward the creation of life.

Whitelaws typology is highly relevant to this research project which involves aspects of both "Hardware" and "Abstract Machines" types. The outward-looking nature of the work and its simplistic embodied mechanism are consistent with the "bottom-up" conception of robotics. Centrally this research is meant to produce a generative system that, although learns from the outside world, is also autonomous and is manifested in a set of computational and formal processes. These processes are not "life as it could be" but are inspired by cognitive models of human thinking.

[16] S. Wilson. *Information Arts: Intersections of Art, Science and Technology*, chapter Research Agendas in Mathematics and Artificial Life. MIT Press, 2002.

This chapter provides a discussion of art practises that intersect with mathematics and artificial life research. At the core is a conception of abstraction—a cultural force that influences both arts and sciences as they attempt to understand and reveal underlying structures, processes, and relationships. This interest in abstraction is facilitated by developments in science and computing technology. Mathematics is not the same as science and technology, it is "the study of structure, order and relation of any kind", and is not inherently physical or empirical.

There are two classes of research agendas in mathematics: (1) Modelling is a process where a model of reality is constructed that balances fidelity and tractability. (2) Pure math is an exploration of elegant geometries, abstract patterns, symmetries, etc.. A-life is an attempt to model life itself, and may be enabled by chaos and nonlinear dynamics. A central research question is: can life exist in a computer simulation? For Edward Shanken and Paul Feyerabend A-life and art are both concerned with the construction of "synthetic creations". Modelling in

science and art are both representational activities. For Jack Burnham the difference between A-life and actual life is stark: A-life is the creation of synthetic biology that represents theories of life. Modelling always involves a choice as to which components are considered essential and which are not. According to Katherine Hayles, this reduction disregards information about life as it is embodied in organisms. Simon Penny sees the practise through a cultural lens and characterizes it as a process that continues science's "totalizing and colonizing" desire to control everything while simultaneously entertaining new perspectives that "probe the impossibility of that goal". Often artists work to become knowledgeable in a particular area of science or technology and then engage in a cultural critique that reveals narratives and concepts that may be invisible to regular practitioners. Nell Tenhaaf believes the interest in modelling links arts and sciences as it manifests an urge to "...develop a symbolic logic and representational system that teases out some kind of order and meaning from a chaotic surround."

While this research project is not conceived of as an A-life project, there are strong overlaps. Modelling is central to this project, both through the use of cognitive models, and also through the creation of a learning system meant to construct a model of the world manifest in a network of concepts. If A-life is the process of modelling the essential structure of life, then this project could be considered in that vein. The elements of consumption and reproduction, deemed as essential, are replaced by embodiment in the real world, perception, memory, and dreaming. As an art-as-research project, the discussion of what the model includes and excludes is paramount. No matter how technically and computationally successful the artwork, it is expected to pale in comparison to the real thing—the model is not the reality.

[17] S. Wilson. *Information Arts: Intersections of Art, Science and Technology*, chapter Algorithmic Art, Art and Mathematics, and Fractals. MIT Press, 2002.

This chapter continues the discussion of mathematics' relation to art, introduced in the previous chapter, with an emphasis on the formal and aesthetic aspects of mathematics applied in media art practises. Algorithms originated in mathematics and are logical sequences meant to lead to particular results. The research of algorithms for image-making purposes is appropriate for artistic enquiry: even traditional arts make use of logical sequences of actions to achieve particular image-making effects. Both the algorithm, and the result, are of interest to the "algorithmic artist". This practise is characterized by the following interests: The artist works with image-making instructions, rather than directly with images themselves. Aesthetic groupings, or "families", of artworks are possible through the adjustment of parameters of a single algorithm. The cultivation of new artistic skills, for example programming, which encourages computer use by artists and allows them to claim programming as an aspect of their artistic practise. The following artists, working in algorithmic arts, are discussed: Roman Verostko, Ken Musgrave, George Legrady and John Maeda (and the media lab "Aesthetics and Computation" group.)

A number of artists are attracted by mathematical methods as it links with "the mysteries of the universe", in short the sublime, and in some cases spirituality. Interests in elegance, beauty and spatial relations (composition) are shared between visual arts and sciences. Efforts at the intersection of art and mathematics aim for "mutual inductance" where both fields are equally enriched by the other. The creation of artworks inspired by pure mathematical forms completely disconnects art from the daily needs of life. Artists working in this area of art and mathematical forms include Stewart Dickson, Brent Collins and Carlo Sequin, Donna Cox, and Brian Evans.

The depth and abstract complexity of mathematical ideas used in artworks could alienate an audience lacking required literacy. These artworks should be "effective" on an immediate visceral level, and later on the conceptual level. Wilson notes a problem with this emphasis on surface form and concept in regards to artists who "are underestimating the importance of that intellectual level; it is part of what differentiates their work and adds additional levels of richness." Some artists include aspects of the intellectual process in exhibitions, showing programs, mathematical proofs and the like. These are often ignored by an audience lacking necessary literacy. Artworks that are strong on the visceral level could facilitate literacy in the viewer by piquing their interest and providing an entry point.

The artistic use of algorithms has a historical precedent including the serialists, minimalism and the conceptualists. The use of algorithms implies a 'systems view' that involves the understanding of structures, processes and interrelationships. Sonia Sheridan argued that "artists could become stronger by explicitly attending to the underlying systems.". Artists working in this area may be divorced from contemporary art as their algorithmic methods are more familiar to science and engineering than they are to contemporary art. The artistic interest in pure mathematical forms could be considered an extension of modernist abstract art that values absolute unity and truth. This conflicts with the postmodern notions of the critical awareness of the social context of art, and the importance of diversity and relativity in contemporary art.

The material from which the artwork produced during this research will be constructed is computational and therefore algorithmic. The importance of the computational process links this research with algorithmic arts. Issues of audience literacy are certainly relevant. One solution is a conception of the artwork that goes beyond the installation and includes papers, presentations and discussions related to the work. This is highly consistent with art-as-research where the artist is expected to participate in the scholarly dissemination of their ideas. Artists may share works with many audiences during conferences and festivals, and even in the media. Key components of this research methodology are knowledge sharing and broad public dissemination.

3 Art-as-Research (Methodology)

[18] K. Busch. Artistic research and the poetics of knowledge. *ART & RESEARCH, A Journal of Ideas, Contexts and Methods*, 2(2), 2009.

Art-research is a broad area with many conflicting and overlapping approaches that are surveyed in this paper. It begins with the assertion that contemporary artistic practises are so theory driven that they are, in fact, already research practises. The nature of knowledge that results from art-research is an "independent form of knowledge" that compliments knowledge attained through scientific methods. The author provides nine classes of art-research that are not entirely delineated, but express nine major approaches.

Art-with-research describes an artistic practise that requires research into theoretical knowledge, and has a long history in visual arts. This practise does not adhere to scientific methods.

Art-about-research is a practise where research is the object of representation, for example painting anatomical studies, scientific instruments, etc., without the use of scientific methods.

In art-as-research, scientific processes and conclusions are used directly in artworks. Research is a central part of the practise that results, not in art objects, but in "critical knowledge". Research is not preliminary, as in art-with-research, or simply a means to an end, but is the "aim of the work itself" and is manifest through symposia, services, publications and interventions.

Art-as-science practise uses rigorously applied scientific methods for the purpose of knowledge transfer. Self-reflection and auto-theorizing are core components. It assumes that art is based on theoretical knowledge and that it can be developed through scientific methods. Art can only be considered a form of knowledge if it "conforms to scientific standards".

Art-about-science is an artistic practise that creates knowledge through the "...examination of scientific knowledge production and its institutions." Scientific methods are not used, as this practise's purpose is the criticism of science.

Art-as-a-different-form-of-knowledge celebrates the difference between artistic practise and scientific methods. The kind of knowledge that results from art references that which cannot be presented or narrated within a "historic structure of knowledge." Art is not just about showing the invisible, but showing to what degree the visible is invisible.

Poetics-of-knowledge involves the direct analysis and criticism of systems of knowledge, including science. In this approach the image and text representations of scientific theory are evaluated as the knowledge and its representation cannot be separated. Theories that evaluate art images can be used to evaluate scientific images.

An art-research hybrid celebrates the open and discursive qualities of artistic practise in order to highlight the restrictions and limitations of the sciences, and to generate theory. Knowledge resulting from art "...cannot be easily brought to a precise point...". The commodification of knowledge means that if theory only serves as a generator of knowledge then contemporary philosophy runs the risk of becoming ornaments attached to art projects.

A practise that demarcates a zone at the margins of disciplines, where art and science can mutually interconnect, opposes the scientification of art. Poorly understood boundaries between disciplines are unknown territories that contain the most urgent problems. Methods, categories and criteria are "in the making" and therefore the object of study is undetermined. Knowledge can flourish here in the "wild" where it does not fit into a particular space or framework. Institutionalized knowledge excludes the unexpected, and so research "...should be inventive to the extreme..." and involve "a fundamental openness to anything that oversteps the framework and conditions of the previously possible, in other words, an openness to experience the unknown, or the impossible."

This research program is explicitly labelled "art-as-research" and involves the direct application of knowledge in cognitive and computer sciences, not as a preliminary stage, but as integral to the art-making process. Criticism of science and systems of knowledge is implicit in this process, where the knowledge being used is not taken to be True, but tried out to see where it leads. Self-reflection and auto-theorizing are core to this research. The grand notion of "a machine that dreams" is deliberately meant to mark out a territory of the "impossible" in order to intersect art and science. Art practise is not usurped by scientific methods, and scientific methods are not ignored. Perhaps only art can initiate a research program into the unexpected where no reasonable hypothesis is possible.

[19] F. Candlin. A proper anxiety? Practice-based PhDs and academic unease. *Working Papers in Art and Design*, 1, 2000.

This paper provides an account of the anxiety experienced by many PhD students in practise-based programs in the United Kingdom. Anxiety is considered endemic in PhD research, but takes on a stronger significance in the case of practise based PhDs, where it may be shared by supervisors, managers and students. Much of this anxiety is due to the artistic identity of students who are often evaluated on entirely scholarly measures. The emphasis on the written component undervalues the artwork, and therefore the artistic practise. It could appear that the broad application of the PhD title among artistic practitioners would undermine the value of the title. The disciplines of art practise and art history may have little overlap in evaluation and competency. This segregation from practise and theory increases the anxiety of students who attempt to bridge both sides.

Strict disciplinary boundaries simplify academic benchmarks and degrees of expected competency among graduates. Authority and mastery are often obtained in well demarcated domains. Art practise is no different in this respect and according to Clement Greenberg should be actively divorced from other disciplines. This has given artists the power and confidence that comes from being within a strong boundary. Artists are also fond of pushing boundaries and redefining art. The infiltration of art in academia could then be considered a natural progression. The methods for evaluating scholarly work do not disappear when artists insert themselves. The artist is left at the intersection of two incompatible disciplines, where they must prove competency in both, which leads to anxiety. This intersection leaves many open questions: For example how should an artwork be evaluated? Should the quality of writing in a practise-based PhD be the same as for a traditional PhD? The institutional pressure to evaluate practise-based work on the written component alone discourages the integration of practise and theory. The effect of practise-based PhD programs may be to mix up boundaries and could encourage trans/multi/cross disciplinary work. The anxiety felt by practitioners in academia is due to the lack of a hard boundary, and the intersection of multiple disciplines. This is also an opportunity to carve out new territory and methods of enquiry. The integration of practise and theory in academia may be inevitable.

This research program is placed at the intersection of art practise, computer science and cognitive science. To gain a PhD level of competency in all these fields is unlikely. This work is about pushing the boundaries of art and fusing ideas from these various disciplines. Perhaps it is due to the research emphasis of this work, but the relation between the artwork and the written component is not an issue. The art is in the written component as much as it is within the code that makes up the non-textual manifestation. Being at the leading edge, a practitioner can look behind and trace the line of development through multiple disciplines. The leading edge is a significant opportunity, despite any "proper anxiety".

[20] A. Haarmann. Artistic Research: A Tool of Cognition Parallel to Philosophy? *Proceedings of the European Society for Aesthetics*, 2:193–208, 2010.

This paper provides an argument for the consideration of visual art practises as research with similarities with philosophy. Three artworks are cited as examples: "Complete Untitled Film Stills" by Cindy Sherman, "Queen" by Candice Breitz, and "Stadt", by A. Haarman (the author). In Cindy Sherman's work, each image in the series elaborates on the research topic—the representation of "female prototypes" in cinema. These representations expose hidden assumptions in cinematic style through three primary methods: (1) The serial construction of multiple images allows the consistencies to be seen. (2) "Trial and error" allow Sherman to find those features of the female prototype used in cinema. (3) Reflection is the visual manifestation of a research practise where the media (photography) reflects the topic (cinema). The deconstruction of moving images into stills allows a "contemplative kind of attention" that is not encouraged by moving images. Sherman is considered a conceptual artist: "[s]he works with ideas and topics and she finds her way of expressing herself artistically in the correlation between form and content." The choice of media results from the topic, as researched in the practise, and the conceptual and questioning aspect of the work is paramount. This emphasis on questioning exposes an underlying interest in process, which is reflected in visual arts and constitute research methods. Artworks are answers to research questions.

The conflict between artistic and scientific knowledge is mirrored in the seventeenth century struggle between scientific and philosophical knowledge. Philosophy was the pinnacle of knowledge and science was considered "nothing else but plain observations". This "knowledge politicking" reflects the social aspect of who decides what research is and what kinds of cognition are appropriate. The individual nature of artistic production may be seen as problematic for scientific reproducibility, and social verification, but art is no less individual than philosophy. Both are embedded in processes of cultural discussion and are therefore not isolated. Artistic and philosophical research both lead to the creation of "monographic" works, where published books and artistic exhibitions are comparable. Haarmann is not convinced that writing is required for art to be considered as research, the example of mathematical notation is cited as a specialized system of representation not easily transcribed as text. Art is simply another example of such a specialized representational system. The application of scientific methodology to art is inappropriate because "...practices, methods, symbolic communication media, and horizons of understanding do differ from one discipline to another." If art depended on methods from other disciplines it would loose its autonomy and simply become a specialization within another discipline. While both philosophy and science are concerned with a study of the nature of the physical world, it is argued that visual art is in a unique position to offer tools and methodology in the study of "...visual culture and the public world of images."

While Sherman uses photography to study aspects of cinema, Breitz studies the "visual and performative aspects of pop culture", through the use of public participation, serialization and media art. In both cases artworks "...can be understood not only as a medium of representation, but as a medium of visual reflection", where the artwork reflects the topic and the context of visual culture. This work embodies a shift from "work-oriented" aesthetics, where the object is of central importance, to "production-oriented" aesthetics, which consider explicitly the ". . . terms of the practises and strategies of the artistic production." The result is a practise where development, experiment, questioning, communication and cooperation are key and the artwork is ". . . responsive to the social reality. . . ". In Haarmann's artwork, the collaborative process is of central concern where the subjects of her film sequences are engaged in an artistic collaboration and responsible for the construction of their own public image. A significant feature of visual arts is how exhibition ". . . carr[ies] visual research into the realm of visibility, making it subject to public debate."

As visual arts are concerned with making images, this visual reflexivity is of central methodological value. Certain ideas can only be expressed through images, and not text or mathematical notation. These images are created ". . . by twisting and turning visual culture by way of visual and performative techniques, thus analyzing its effects." In short the ideal methodology for studying visual culture is the participation in visual culture through the construction of images. This not only reflects visual culture but also "enlarges" it. Public intervention, though exhibition and collaboration, give such work an ". . . interventionist character, interfering with the aesthetics of everyday life that it researches and with the world of public images." Key methods in visual arts research are serial elaboration, image generation, cooperation and intervention. Visual artistic research methods ". . . treat and present different cultural realities and make them debatable and perceivable."

The conception of art-as-research presented in this paper appears to only apply to visual arts—where the image itself is the form of enquiry, and the topic is linked with visual culture at large. The examples provided exist for the purpose of exposing hidden structures of meaning to which we have become habituated. The argument could be extended to include electronic media art practise. If visual arts is in enquiry of visual culture for the purpose of exposure through public exhibition of visual artifacts, then an art-computation practise could be an enquiry manifested in a computational artwork also meant for public exhibition. The framework could be applied to this research as follows: an artistic enquiry of cognitive science (dreams, perception and development) and computational models of mind (artificial Intelligence, agents and machine learning) involving the construction of a computational artwork that implements these models that engages in a broader public discourse through the dissertation and exhibition. As discourse is central to this research, it is appropriate to look beyond just the art object as a site of enquiry.

[21] N.L. Holt. Representation, legitimation, and autoethnography: An autoethnographic writing story. *International Journal of Qualitative Methods*, 2(1):18, 2008.

Autoethnography is a qualitative research method, situated in the postmodern tradition, that raises doubts regarding the "...privilege of any method of obtaining authoritative knowledge about the social world." The methodology involves the writing of highly personalized accounts where "...authors draw on their own experiences to extend understanding of a particular discipline or culture." These texts often feature dialogue, emotion and self-consciousness. This paper discusses the crises of legitimation and representation in autoethnography through the lens of an attempt to publish such a work in a journal. The purpose of autoethnography is to confront dominant forms of research knowledge (representation), in marginal spaces, through self-reflection. In creating the autoethnographic account, the author draws on a reflective framework, applies a methodological tradition, and links results back to the object of study. The author's attempt to publish an autoethnographic account of a pedagogical experience is analyzed.

The feedback from seven reviewers and four editors are integrated into two recreated discourses that reflect the review process as a whole. Reviewers fall into two broad categories: (A) those that valued ethnography, but were unsure how to evaluate it, and (B) those who do not consider autoethnography a valid research practise. 'A' reviewers tend to fault the work on its lack of rigour, while 'B' reviewers fault the entire methodological approach. An autoethnographic work can be validated "...if it evokes in the reader a feeling that the experience is authentic, believable, and possible." Richardson provides five evaluative factors: The work should make a substantive contribution, have aesthetic merit, be reflexive, impactful and genuinely expresses a reality.

A hybrid autoethnographic approach is considered for this research work. As an artistic practise much can be gleamed from the process of creativity, and as such the practise is logged. Autoethnographic writing requires an artistic and emotional text that may not be appropriate for this project. At its core, autoethnography is a celebration of the self, and requires trust in the self through a research process. This is a trait in common with artistic practise, where the work is constantly being reevaluated in relation to the self.

[22] M. Mateas. Expressive AI: A Hybrid Art and Science Practice. *Leonardo*, 34(2):147–153, 2001.

Expressive AI is an approach to art making that combines concerns from AI and cultural practise. In this paper, Mateas describes three works of art, and provides a description of the expressive AI approach. "Subjective Avatars" actively manipulate the viewer's perception of a virtual world through the use of a "personality model" that maintains an emotional state and narrative context that is independent of the viewer. "Office Plant #1" fills the emotional niche of a real plant in a high-tech office. The robot is plant-like and responds to the mood of the email activity of the owner. Mood is determined through text classification, and is mapped to smooth and slow movements and a sound-scape. The mapping between email mood and the artwork's behaviour is accomplished using a fuzzy cognitive map. "Terminal Time" is an interactive cinematic experience that expresses ideologically biased documentary histories. The audience manifests its will through an applause meter, which influences a goal-tree formulation of ideology that is used to generate a spoken synthesized text and select video clips from a media library.

Mateas describes the differences between the two major AI camps: GOFAI is characterized by an emphasis on symbolic logic, problem solving, depth over breadth, generality and planning. Interactionist AI is concerned with embodiment, breadth over depth, and the use of world as model. In general AI, the goal is to build systems that exhibit intelligent behaviour. This is accomplished when a well defined task is "demonstrably accomplish[ed]", and the result is expected to have broad utility. These interests conflict with artistic practise where artifacts "convey complex meanings, often layering meanings, playing with ambiguities, and exploring the liminal region between opaque mystery and interpretability." Art "... is a stance or viewpoint from which all of AI is reconstructed."

For Mateas, the purpose of an artwork is to communicate with the viewer. This corresponds to the "conversation metaphor" where the artwork itself is considered a communication channel. AI is concerned with building machines that *are* intelligent, while art making is about the construction of machines that *appear* intelligent. In art, the internal structure of the artwork is accidental to its purpose of communication. In AI, any meaning-making in the work is accidental as its purpose is the objective modelling of intelligent processes. Expressive AI is concerned simultaneously with internal structure and communicative potential, where both are developed in tandem. As art is a communicative medium "[a] focus on the machine is alien to current electronic media practice", as it only exists to support this communication. In extreme cases, this lack of interest in the internal structure of the machine results in situations where "...the machine does not actually do what is indicated in the concept of the piece." Expressive AI allows a "a depth, a richness, a sophistication" that cannot be achieved without simultaneous concerns in internal structure and communication. Expressive AI is not the "mere" application of AI methods. AI can be applied in "toy" cases where they have little AI value, but can be significant to the viewer. Expressive AI practitioners reflect on the relations between structure and communication. Cultural theory is a useful tool for unpacking the hidden assumptions in the background of traditional AI interpretations. "Expressive AI changes the focus from an AI system as a thing in itself (presumably demonstrating some essential feature of intelligence), to the communication between author and audience."

The artistic practise in which this research is situated does not consider the purpose of art communication or expression. It is concerned with both cultural practise and the internal structure of the systems it produces. As such it could be considered an expressive AI practise.

The purpose of this research is the exploration of both cognitive models and computational methods through an artistic practise. As such the internal mechanisms are paramount, and the aspect of communication secondary. This artistic practise aims to resolve the same problem highlighted in this paper—the disconnect between the reality of an artwork, and its conceptual description.

[23] D. A. Schön. *The Reflective Practitioner: How Professionals Think in Action*, pages 3–71. Basic Books, 1983.

Chapters 1 and 2 in this book provide a description of the "crisis of confidence", a description of "Technical Rationality" and presents Schön's proposal for reflection-in-action. The "crisis in confidence" occurs in a context where the number of professions is exploding and they experience difficulty maintaining the demands put on them by society. A professional is an individual who has a specific set of skills and operates in a well defined domain. The crisis can be summarized as a mismatch between knowledge in traditional practice and the complex, uncertain, unstable, unique and conflicting parameters of reality.

Technical Rationality describes professional activity as instrumental problem solving which is made rigorous through the application of scientific theory and method: the application of general principals to solve specific real-world problems. Practice is subservient to research because research provides the basic and applied science from which problem solving methods are derived. Practice provides real world examples for scientific study, and tests the utility of scientific research.

A fundamental aspect of reflection-in-action is "problem setting". In problem solving, the context of the problem is often ignored, which may result in the practitioner missing out on the big picture. Problem setting decomposes a problem into components, and explicitly situates the problem in a broader context. During practise, actions, recognitions and judgements can spontaneously form without conscious thought. The practitioner may not remember learning this tacit knowledge because it is slowly internalized through practise. Components of practice can be combined, recombined and varied in an improvised fashion. Reflection-in-action can combat over-learning by allowing the practitioner to explicitly reflect on these implicit choices and therefore see the problem in a new light.

Schön's description of a self-reflective practice is highly congruent with artistic practice where the artist is constantly reconsidering and reflecting on the current state of the project, its reasons for being, and the context of production. In this research a journal will provide documentation of process that can be reflected upon, both during and after, project development.

[24] B. Steinheider and G. Legrady. Interdisciplinary collaboration in digital media arts: A psychological perspective on the production process. *Leonardo*, 37(4):315–321, 2004.

The process that lead to "Pockets Full of Memories" (PFOM) is an example of collaboration between a number specialists in various disciplines in different geographical locations and speaking multiple languages. PFOM is a complex media art installation that combines art, technology, science and design. The increasing complexity of media art projects leads to a movement from individual to group collaboration. A core requirement in these collaborations is the development of shared and multifaceted understanding of the "problem". Knowledge sharing is the means by which the team develops consensus regarding project goals, the understanding of the various members' expertise, and the meta-knowledge that connects those areas. Aspects of communication that foster positive communication are openness, trust, a willingness to compromise, sympathy, spatial proximity and abilities in technical communication. Features of coordination that facilitate collaboration include systematic project and time management, and stability in team composition. A central area of conflict in PFOM is rooted in the two major development styles: the engineers and scientists tend to favour a top-down approach, where the project is developed through a series of specifications, while the creative art and design group prefer to experiment and materialize ideas in order to evaluate them and make further developments.

Although his paper describes a collaborative process, the issues with interdisciplinary work are shared with this research project that intersects art, cognitive and computer sciences. This project requires regular discussion with specialists in these disciplines throughout the research process, and therefore requires similar approaches to knowledge sharing. The creative production described in this paper is framed as a "problem solving" process, which is a problematic view of artistic practise where the "problem", if it even exists, could change during development.

[25] J. Strehovec. New media art as research: art-making beyond the autonomy of art and aesthetics. *Technoetic Arts: a Journal of Speculative Research*, 6(3):233–250, 2009.

This paper describes the shift of contemporary artistic practise, in particular new media art, from an autonomous discipline tasked to create aesthetic objects to an "art service" where fluid and temporary processes integrate science, technology and life. While traditional aesthetics deal with the specifics of the artwork, its genre, form, creativity, author, and value, new media art is seen in terms of software, process, experience, service and interface.

In new media, the autonomy of art is rejected and integrated with social reality. Artists challenge the institution by conducing research, framing concepts, and creating new paradigms that go beyond the limits of their field. Cultural innovations increasingly result from new media practitioners and not painters, filmmakers or writers. Artworks are constructed for the purpose of knowing what art is. Research, rather than representation, becomes the key function of new media art. The author cites the work of "Critical Art Ensemble" (CAE) whose public performances make initiatives in science immediate and concrete for the public. Art stands up as an example of research that is different than scientific research and is still "oriented towards knowledge production." Components of artistic research practise include "... activism, hybridization, recombination, (re)mixing, repurposing and researching." The shift to art-as-research is evidenced by artists' increasing preference for terms like "researching" over "creating" or "forming". Research is not the only function of art but the "... play of the scientific, the artistic, the technical, the political and the conceptual ... " are also central. The shift from artifact to service is mirrored in other areas: for example, the shift from material to immaterial wealth through the computerization of the economy. The "art service" is a process oriented, task driven, systematic approach to a particular "problem", where the results are not an artifact but a solution meant for inclusion in the everyday. When considering art in this sense, art-making is challenged to simultaneously participate in other fields, and to therefore abandon "art's pure autonomy."

The approach to new media art that is proposed above is highly congruent with the art-as-practise proposed for this research. Particular overlapping ideas are the importance of the integration of art and everyday life, the use of art to share scientific knowledge, and the consideration of the artwork not as an object, but a process that exists for a particular purpose. In the background of this research is the use of only free and open-source software, which is an intentionally political act, and is integrated into the very artistic process.

[26] G. Sullivan. *Art Practice As Research: Inquiry In The Visual Arts*, chapter Practice as Theory. Sage Publications Inc, 2005.

A broad overview of various practises in visual arts research are discussed in this chapter. In particular a framework for visual arts research projects is presented, as well as a number of links with empiricist, interpretive and critical research practises. In "practice as theory" disciplinary boundaries are considered bridges and not barriers where practice and theory collapse through the efforts of artist-theorists. Sullivan points to the increase of transdisciplinary collaborations as an example of this, where research is guided by questions, issues and abstractions where knowledge creates and critiques human experience.

The visual arts framework constitutes a practice that is characterized by being dynamic, flexible and revelatory, while traditional research is linear, interpretive and confirmatory, and qualitative research cyclical, emergent and discovery oriented. The framework requires that the research be systematic, rigorous and imaginative. Practice-as-theory involves research

through three manifestations: (1) Visualization is a creative, critical and clinical method of constructing images to represent objects, data, texts and ideas. (2) Artworks are presented in exhibitions or performances where artists, curators, art writers, academics, educators, institutions are communities intersect and allow an enquiry "at the public interface". This collaborative and public-oriented context allows the collection of "credible" data. (3) Exegesis is the traces of a visual practise that are separate from the artwork; it serves preliminary purposes, records in progress activity, and displays outcomes.

Arts research practise can appropriate methods in the empiricist tradition where research is a quest for new knowledge that is shaped by questioning what is known and offering new conceptions of "what is, what might be, and what ought to be". Strategies include reviews and reports, empirical methods, such as collecting and analyzing data, and design and simulation, where the details are abstracted in order to present a project at a high-level. According to the interpretive tradition "...meanings are made rather than found", and are constructed individually and socially. Visual arts research in this vein is reflective, contextual and reflexive. It depends on dialogue with the broader community by bringing personal accounts into the public domain. Visual arts research has a long history of criticism, which is considered a global/local and theoretical/practical collaborative process that involves the systematic and rigorous comparison and contrast of positions, assumptions and arguments.

Sullivan provides a extremely broad survey of intersections between artistic practises and research traditions. Specific methods and techniques for conducing research through a visual arts practise are discussed. This research is not primarily a visual arts practise. Image-making is certainly involved, but the emphasis is not on the construction of images for research purposes, which is essentially the bulk of Sullivan's discussion. He describes accepted artistic practises and inserts them into a tradition of research, which implies an inherent lack of validity and rigour in artistic practise itself.

[27] S. Wilson. Artificial intelligence research as art. *Stanford Electronic Humanities Review*, 4(2), 1995.

In this paper, Wilson discusses potential contributions that can be made by practitioners in arts and humanities in the development of AI systems. These contributions are situated in Wilson's conception of "Information Arts", characterized by three possible stances that artists can take in regards to contemporary technology: (1) Deconstruction, as an art practise, involves the postmodern and structuralist analysis of texts, narratives and representations that underlie contemporary technology. (2) Artists may continue a modernist conception of art as a "specialized art discourse with claims to universal aesthetic truth" and/or (3) the invention or elaboration of new technologies and their cultural possibilities. Information Arts aims for an integration of art and science where artists develop proficiency in science and technology in order to contribute to both fields.

Wilson describes his AI related practise through a number of artworks. Wilson's exploration of AI involved learning research agendas, accomplishments and unresolved problems, taking courses, learning programming and corresponding with researchers. He used this contact to expose areas of research that seemed undeveloped, and produced art installations that focused on issues that arose. The artworks discussed involve themes including: (1) telecommunications and telematics, (2) interactivity and the audience, (3) hypermedia and information structure, and (4) artificial characterization and intelligence. "Is Anyone There?" is a system that uses the public telephones to engage the casual passersby in a conversation with an artificial entity. The project was manifest in two stages, one that involved only public intervention, and a second that involved a gallery installation in combination with public intervention. "Excursions in Emotional Hyperspace" is an interactive installation consisting of four characters manifest as mannequins. By walking between them, the viewer hears a personal narrative, told from each mannequin's perspective, and in reference to the previous speaker. "Time Entity" is an alien character that is bound to, and obsessed with, the passage of time. The entity develops through stages and appears to be aware of its own mortality. The passage of time can be manipulated by the audience through a tactile interface. "Demon Seed" is a set of robots that are meant to express our contradictory fear and attraction to robots and new technology. Each robot appears as a demon as mythologized by various cultures, responds to a tactile interface, and vocalizes through "demonized" recordings of the viewer's voice.

The central argument in this paper is that humanities and arts have a lot of knowledge that can benefit AI research. AI systems are build upon many choices, but not all of these are purely technical. These non-technical areas are where contributions from other fields can be made. Wilson identifies three aspects of AI research which require contributions from other fields. The physical basis on which AIs communicate with users is an area that requires exploration. The keyboard/CRT may not be appropriate interfaces and Wilson stresses the possibilities of tactile interfaces. Non-human-centric methods of communication may be lead to subtle and animal-inspired approaches. The limited domains in which AIs are typically applied are problematic as variation is an essential aspect of human communication. The application of AIs in general contexts require other disciplines. AI systems should adapt to us, rather than the other way around. The development of functional characters and narratives in art and literature could be the basis of AI's models of themselves, the world, their relation to others and their partners.

This research is expected to enrich AI, computational and cognitive domains through the construction of artworks that intersect with those fields and provide a cultural perspective and criticism of the models and concepts. Wilson's conception that artists should participate in science and technology is valid and significant. The conceptions and theories developed through art have a valuable bearing on technological development, which is considered a cultural endeavour.

4 Brain Science and Computer Vision

[28] R. M. Benca, W. H. Obermeyer, C. L. Larson, B. Yun, I. Dolski, K. D. Kleist, S. M. Weber, and R. J. Davidson. EEG alpha power and alpha power asymmetry in sleep and wakefulness. *Psychophysiology*, 36(04):430–436, 1999.

A study of the asymmetry of cortical activation between brain hemispheres is described in this paper. The activation of hemispheres is inversely proportional to "alpha band power" ("alpha power" hereafter)—the measure used in this study. Alpha power is shown to be stable over time and exhibits "excellent internal consistency and reliability" (Benca et al. quoting Tomarken et al.). It is established that the dominance of hemispheres during waking, as measured by alpha power, is associated with "emotional reactivity". Subjects that show large alpha power asymmetry are associated with particular traits: Those with low alpha power (high activation) of the left frontal area tend to have more positive dispositions. Those with high alpha power in the right frontal area tend to respond more negatively to stimuli "... designed to provoke negative effect". A recent study has shown that differences in frontal asymmetry predict the emotional content of dreams: subjects with a negative disposition, corresponding to right dominance, tend to have dreams with negative emotional content.

Results show correlations between waking and sleep EEG asymmetry, indicating that dominance during waking caries over into sleep. Additionally, the correlation between waking and sleep asymmetry is greatest in REM sleep. While there is tendency toward stability in frontal and temporal asymmetry across sleep stages, the study indicates a shift of alpha power toward the right frontal hemisphere in both REM and NREM sleep. This could indicate that dreams are more correlated with high activation of the left hemisphere, and is supported by fMRI studies. An earlier EEG study (Goldstein et al.) found that REM is more correlated with higher amplitude in the left hemisphere, while NREM is more correlated with higher amplitude of the right hemisphere. It has been shown that during sleep onset, the alpha power in the frontal areas increase (decrease of activation), while the posterior areas decrease (increase of activation). A final note is that the correlation between alpha power and cortical activation has not been proven to hold across sleep stages.

The traits corresponding to hemispherical dominance confirm those in Sperry's account. In short, right hemispherical dominance may be correlated with creativity, or at least aspects of creativity. On an intuitive level, the relation between dreams and creativity is obvious, but are there neurological overlaps between dreaming and creative cognition? The shift of activity to the left hemisphere during dreaming is counter-intuitive as it implies a correlation between dreams and the analytical side. The work of Goldstein et al. shows a correlation between REM sleep (where dreams tend to be bizarre) and the "creative" right hemisphere, and the correlation between NREM sleep (where dreams are shorter and more thought-like) with the "analytic" left hemisphere. If the right hemisphere and dreaming are correlated with origination, and the left analytic evaluation, then dreaming should be dominate towards the right. On the other hand, if dreaming is the random activation of perceptual mechanisms that are "made sense of" by the brain, then this could explain the shift to the analytic side. Of course as the correlation between cortical activation and alpha-power is not confirmed in all sleep stages, the link between the left hemisphere and dreaming could be incorrect.

[29] R. Borisyuk, Y. Kazanovich, D. Chik, V. Tikhanoff, and A. Cangelosi. A neural model of selective attention and object segmentation in the visual scene: An approach based on partial synchronization and star-like architecture of connections. *Neural Networks*, 22(5-6):707–719, 2009.

A system making use of a Spiking Neural Network (SNN) for object selection and attention is described in this paper. The SNN is a variation of an artificial neural network (ANN) where neurons "fire" at a frequency that is proportional to the degree of stimulation. This is simplified in an ANN where the value of a neuron represents the rate at which its biological counterpart would fire. The task of object segmentation requires context, previous experience and goals. According to the Temporal Correlation Hypothesis, attention is correlated with synchrony of neurons. The system is composed of three modules: Module A selects a region of interest, from test images, that is further processed in Modules B and C.

Module A is the selective attention module. It is a network of neurons implemented according to the Hodgkin-Huxley model. The network contains central (C) and peripheral (P) neurons. Image features are fed into the P neurons and correspond to areas within the image, while the C neurons correspond to the "Central Executive" and control attention. Attention is considered synchrony between C and P neurons. The C neurons inhibit incoming connections from P neurons that correspond to areas of the image not attended to. In test images, consisting of four coloured balls, the background is hard-coded so that it stimulates P neurons to a very small degree, when compared to objects. Local excitatory connections between P neurons allows a spreading of activation, at a particular rate. This allows similar colour regions to be represented by similar firing rates of the corresponding P neurons. Module B implements a traditional method, using Gabor filters and derivative analysis at multiple scales, to extract object contours from the test image. Module C uses data from Module B and A in order to find edge boundaries and segment objects. It is composed of a star-like architecture of coupled phase-oscillators. It suppresses noise and spurious inclusions through the spreading of synchronous activation, similar to that in Module A, and is constrained by contours calculated by Module B.

The SNN is an attractive approach to AI because its is more consistent with neuron anatomy and includes well known temporal properties that are biologically based. Unfortunately they are computationally inefficient when compared to ANNs, which makes real-time processing unlikely, and their temporal properties moot. The model proposed in this paper depends on traditional AI techniques for segmentation, and the attentional mechanism is simplistic, and seems to centrally depend on the hard-coded saliency between objects and background. It is unclear how attention relates to dreaming, as it is presumed that all components of dream experience are equally attended to. Dreams are not intentional in the traditional sense, which is paramount to attention.

[30] F. Crick and G. Mitchison. The function of dream sleep. *Nature*, 304(5922):111–114, 1983.

This paper describes an early theory of the function of dreams as the "clean-up" of spurious associations in the cortex. The widespread occurrence of REM sleep in mammals and birds points to a significant function of dreams. The authors describe the major features of biological neural networks. Patterns are stored such that they are distributed, robust, and superimposed. They can complete partial input patterns (are content addressable) and can perform classification. In simulations it has been shown that networks can "overload" when patterns are superimposed at too great a degree. This may result in bizarre associations (fantasy), outputs unrelated to inputs (obsession) and spontaneous outputs (hallucination). It is expected that the creation of associations, in biological neural networks, would lead to "parasitic" connections that do not correspond to useful learning. The authors propose that REM sleep may have the function of removing these associations by: (1) turning off inputs and outputs, (2) feeding the network with random values and (3) dampening the activated spurious nodes.

Spurious nodes are selected because they tend to be activated by random patterns, rather than by coherent stimuli. Random activation comes in the form of PGO (pontine-geniculate-occipital) waves originating in the brain-stem. The value of a reverse-learning processes was confirmed in simulation by Hopfield, where it was shown to "equalize the accessibility" of stored memories and suppress spurious ones. Dreams are stored in short-term memory, and consequently are not remembered unless the subject awakes during the dream, and even then often partially. Since dreams are the manifestation of spurious memories, recurrent dreams are explained because the remembering of dream content, due to stressful emotional content, switches reverse learning into pattern reinforcement where spurious associations are more deeply engrained. They even suggest that remembering dreams should perhaps not be encouraged, as it could interfere with this "clean-up" process. The authors have no significant argument against the lack of ill effects in subjects taking drugs that disable REM sleep. If the theory is correct then subjects lacking REM sleep should experience some symptoms, including obsessions and/or hallucinations.

The notion of "clean-up" is similar to the concept of the garbage collector in OOP. As "reverse-learning" it could also be considered a negative reinforcement for certain patterns, which could still be considered an aspect of learning, and likely how it was implemented in Hopfield simulations. This paper is the foundation of major features of Hobson's AIM model, including the function of random PGO waves and the disconnection of the body. The argument that random activations would likely activate spurious associations appears weak, as its not clear what would disable inter-neuron activation spreading from spurious to valuable associations.

[31] M. A. Goodale and D. A. Westwood. An evolving view of duplex vision: separate but interacting cortical pathways for perception and action. *Current Opinion in Neurobiology*, 14(2):203–211, 2004.

This paper describes the function of the two visual pathways that connect the occipital-parietal and the occipital-temporal lobes. According to Ungerleider and Mishkin, the "dorsal stream" and the "ventral stream" deal with spatial and object centric processing, respectively. Goodale et al. propose an alternative interpretation of lesion studies of these areas: The dorsal stream provides the visual control of action, while the ventral stream provides visual perception of the world. Both pathways deal with both object and spatial processing.

Goodale et al. make a distinction between two different ways of seeing: one is conscious and allows the perception of visual objects while the other is unconscious and allows the integration of motor and visual routines, for example reaching for a red ball. D.F. is a subject with a brain injury in the ventral steam. She is cannot see, or recognize, objects but she unconsciously changes the size of her hand to match the size of objects she can grasp. As the visual-motor action of reaching for a ball requires both the recognition of the ball, and the motor integration of the ball, the dorsal and ventral streams are not independent, but interact to some degree.

This paper provides an alternative to Ungerleider and Mishkin's interpretation of lesions studies in dorsal and ventral visual streams. The kind of visual processing that "Dreaming Machine" will do is likely specific to the ventral stream, as no motor appendages are planned for the installation. There does appear to be evidence that the dorsal stream may be correlated

with eye movements. The choice as to whether "Dreaming Machine" will have a static or moving camera is not decided.

[32] K. S. Graham, M. D. Barense, and A. C. H. Lee. Going beyond LTM in the MTL: A synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory and perception. *Neuropsychologia*, 48(4):831–853, 2010.

The medial temporal lobe (MTL) has long since been correlated with memory, while damage to the hippocampus has been correlated with deficiencies with spatial navigation. Recent studies show that the MTL is correlated not just with memory but also with perception. According to the Emergent Memory Account (EMA), memory involves the interplay of multiple systems throughout the brain, not just in the MTL. It predicts that memory deficits resulting from MTL damage are due to the damage of memory representations that are shared between perception and memory.

It is clear that the MTL is critical to certain aspects of human memory. The hypothesis that the MTL is the central memory area is not supported by studies that show that those with "conscious declarative memory deficit" can still learn via unconscious non-declarative memory. It was assumed that the parahippocampus, the hippocampus, perirhinal and entorhinal cortecies constitute a single system for declarative memory. Recent work shows a functional specialization within the MTL. The hippocampus, the fornix, the mamillary bodies and the anterior thalamus are all critical for encoding and recall of episodic memory, but not correlated with familiarity-based item memory (memory of objects). According to Relational Memory Theory the hippocampus is involved in the rapid learning of associations between items and their contexts. While the hippocampus is related to the temporal and spatial organization of items, the perirhinal cortex is correlated with object-centric features. According to the Binding of Item and Context (BIC) theory, the perirhinal cortex handles objects while the posterior parahippocampal cortex handles the storage of contexts, both spatial and non-spacial. Additionally, the hippocampus binds spaces and objects. According to this theory, the MTL is correlated only with memory and not perception. The perirhinal cortex is the apex of the ventral stream and is used to resolve feature ambiguity by integrating many visual features. The hippocampus is correlated with scene recognition, and representations are implicated in both long and short term memory. According to the EMA hypothesis perception and memory depend on shared representations that occur in both the hippocampus and the perirhinal cortex.

Early work on a prototype perception system for "Dreaming Machine #3" (DM3), conducted during the metacreation class, centred on a separation between object-centric and spatial visual processing. This separation was inspired by the "where vs what" theory, proposed by Ungerleider and Mishkin, which states that the ventral visual stream is object-centric while the dorsal stream is spatial-centric. The theory is not well supported by contemporary evidence. The contemporary manifestation of object and spatial processing, discussed in this paper, is manifested in the perirhinal cortex and the hippocampus, respectively. This separation is a hint to the architecture of the visual system and should be integrated into DM3's visual system. Additionally, poor infant visual acuity and Mandler's theory of conceptual development point to high importance of spatial, over object-centric, processing.

[33] J. A. Hobson. REM sleep and dreaming: towards a theory of protoconsciousness. *Nature Reviews Neuroscience*, 10(11):803–813, 2009.

Hobson proposes that dreams function as a "virtual reality" for the protoconsciousness to develop during sleep. Hobson distinguishes between two types of consciousness: Primary consciousness largely involves perception and emotion while secondary consciousness depends on language, self-awareness and involves abstract thinking and volition. From studies of the content of dreams, Hobson concludes that dream consciousness is similar to primary consciousness because intense perceptual and emotional content is common, and experiences tend to be incoherent, fantastic, bizarre and contradictory.

The paper describes the three objectively observable sleep-wake states: Waking, Rapid Eye Movement (REM) sleep and Non-REM (NREM) sleep. The sleep stage correlates with activity of particular neuron groups: Aminergic inhibitory neurons are activated during waking and inhibited during REM sleep, while cholinergic excitory neurons are inhibited during waking and activated during REM sleep. During REM sleep, the brain is minimally inhibited, which results in broad activation of brain regions.

Random signals thought to cause dream experiences are found in the pontine brainstem, the lateral geniculate and the occipital cortex (PGO). These signals are treated by the brain as external and provide the virtual reality available to sleep consciousness. This PGO activation of the visual system originates in the low-level motor system, deep in the brain. Hobson proposes that this virtual reality is a safe harbour for the development of the protoself. The protoself forms in order to take responsibility of the automatic acts that are executed by cognitive processes. For Hobson volition in dreams is an illusion, just as consciousness is an illusion during waking.

The AIM model consists of three primary processes: activation, input/output gating and modulation. During REM and waking the brain is highly activated, during NREM sleep the brain is less activated. During REM sleep, the reticular activating system disconnects (gates) the forebrain from the sensorimotor system (input signals from the body are not easily transmitted during sleep). The inhibition of aminergic neurons during REM is expected to be the cause of memory deficit. During REM, a number of neurotransmitters that are active during waking are switched off.

Hobson provides a case for a "bottom-up" model of dreaming, where a perceptual system is activated by random signals, which are interpreted by the perception system. This aligns with the consideration of dreams as similar to primary consciousness, as the perceptual system is activated directly from PGO signals. The conception of dreaming as key in the development of the protoself and protoconsciousness could lead to computational and developmental conceptions of self. The paper provides many details concerning the content of dreams and their possible neurological basis. The AIM model itself is not well suited to this research because the random activation of a perceptual system is similar to the system implemented in the early Dreaming Machine projects. Nir and Tononi provide an alternative conception of dreaming that is top-down: dreaming originates in the activation of a high-level conceptual system which generates perceptual experiences. This position is supported by the correlation between children's dreams and their imaginative (mental imagery) abilities.

[34] J. Smith A. Cohen M. Clark L. Tucker, A. Thompson and M. Barrett. Why do we dream? Documentary Television Program, Feburary 2009. BBX Horizon Documentary.

This documentary video discusses a number of theories of dreams. Overall, there is consensus that dreams have purpose and may be meaningful. There are differences in emotional experience between REM and NREM dreams. The former tends to show an increase of "negative regard", while the latter tends toward "positive regard". This difference in emotion indicates that the amygdala is involved in dreaming. Mark Solms found that damage to the parietal lobe can lead to subjects lacking dreams. This region is correlated with the integration sense data and body representation. Patients with this particular injury are often awakened by the onset of REM sleep. For this researcher, dreams are considered "motivated search". Antonia Zardra studies dream content in individuals. If dreaming is not random, and is more like imagination, then dreams are meaningful because they are expressions of the world-view of the dreamer. Robert Stickgold has found that dreaming can improve some skills. He sees dreams as a free-associative way of moving through memory, linking them more freely. According to this notion, dreams could create meaning by forming new links between memories. Anti Revonsuo, who studies children's nightmares, thinks that dreams are "biologically programmed". Nightmares could be rehearsals for stressful events, preparing individuals for future stress. In adults, dreams often pertain to modern threats. A key statement he made is that: "We are dreaming all the time, its just that our dreams are shaped by our perceptions when awake, and therefore constrained."

This broad overview of dream research provides a characterization that will influence this research project. As stated before, this project subscribes to a conception of dreaming as similar to imagination. As such dreams are considered meaningful probes into the conceptual structure (world-view) of an individual. The emotional content of dreams is not directly addressed due to the problem of machines having emotion. A conception of emotion as abstraction of biological needs is a reasonable starting point, but it is not clear what these needs would be. One possibility is the occurrence of faces in visual images, which could serve as an emotional basis. The differences between NREM and REM sleep are not clear, and it is not likely that this research will have a machine that distinguishes between REM and NREM sleep. It appears that the "negative regard" associated with REM sleep, correlated with the right hemisphere, corresponds to traits associated with right-dominate subjects, and vice versa. If REM and NREM sleep are strongly associated with right and left hemispheres, then that is sufficient reason for the artwork to distinguish between those sleep stages. Robert Stickgold and Anti Revonsuo appear to subscribe to Hobson's hypothesis that dreams are a "virtual reality" in which dangerous situations can be experienced without physical harm and allow the practise of skills. The link between visualization and performance is well established, which implies a connection between dreams and imagination (mental imagery). The link between waking experience and dream content, as due to latent activation in the brain, is implied by Revonsuo's quote above. Perhaps we are always dreaming (and imagining) it is simply that perception mechanisms dominate waking experience, while dreams result from the continuing activation of the brain that is no longer constrained by perception.

[35] M. Mishkin, L. G. Ungerleider, and K. A. Macko. Object vision and spatial vision: Two cortical pathways. *Trends in neurosciences*, 6:414–417, 1983.

This seminal paper describes a separation of brain regions specific to object and spatial centric visual processing. There are two functionally separate branches in the visual system: the "ventral stream" passes from the occipital to the temporal lobe, and is correlated with object centric processing, while the "dorsal stream" passes from occipital to the parietal lobe, and is correlated with spatial centric processing.

The ventral stream allows the extraction of qualities of stimulus that are independent of location in the visual field. This pathway is crucial for object identification and is sensitive to qualities such as size, colour, texture and shape. In lesion studies, the removal of this area in monkeys leads to "...impairment both in the retention of visual discrimination habits acquired prior to surgery and in the postoperative acquisition of new ones." Subjects are left unable to distinguish objects that have been learned, and unable to learn to distinguish between new objects.

Lesioning of the dorsal stream, in monkeys, leads to impairment of the ability to locate objects in the visual field, but not the ability to recognize them. The conclusion is that the dorsal stream processes location information independent of object-centric features. Impairment correlates with lesion size and not lesion location in the dorsal stream. Whereas the fovea is highly connected to the ventral stream, the dorsal stream is highly connected to both foveal and peripheral areas.

The functional separation of object and spatial centric processing is significant to this project as the "Dreaming Machine" is expected to perceive the world in a way similar to a human. Although the theory is not strongly supported by more recent evidence, the concept of a functional separation of spatial and object processing could still be valid. A prototype system has been constructed that does basic computer vision using two parallel pathways for object and location centric processing. According to Mandler, object location is more significant than object features in infant object recognition and conceptual development. This is partly because infants have poor visual acuity and are unable to distinguish objects based on object-centric features alone. If infant conceptual development is based on spatial, and not object, visual processing then "Dreaming Machine" should reflect this. This conception solves problems inherent in object recognition using only object features, for example the

early concept of animate vs inanimate objects which cannot be easily distinguished based on object-centric features.

[36] E. A. Murray, T. J. Bussey, and L. M. Saksida. Visual Perception and Memory: A New View of Medial Temporal Lobe Function in Primates and Rodents. *Annual review of neuroscience*, 30:99–122, 2007.

This paper continues the evolving view of the role of the medial temporal lobe (MTL) in perception and memory and discusses a separation of object-centric and spatial processing. The hippocampus is correlated with memory and perception of places, paths, scenes and layouts, while the perirhinal cortex is associated with objects, and the contents of scenes.

Object-centric vision is thought to be carried out by the "object-analyzer" pathway, which consists of the striate, the primary vision cortex (PVC), the secondary visual cortex (SVC) and the inferior temporal lobe. This pathway integrates multi-modal information from other brain regions and "construct[s] higher-order visual and multi-modal representations of objects." Studies in non-human primates show that damage to the perirhinal and entorhinal cortex lead to the same memory deficits. These subjects also show impairment in perception with small and non-coloured stimuli. These deficits tend to increase as the number of stimuli in the test set increase. The hypothesis is that the perirhinal cortex serves to store representations of complex conjunctions of many features. In human studies these results are largely mirrored—perception and memory are both impaired with MTL damage. It has been shown that the perirhinal cortex shows higher activation when presented with a novel object than when exposed to a known object.

There is evidence of a functional separation in the MTL, as damage to the hippocampus is correlated with very little or no impairment during object recognition tests. In non-human primate studies, the increase of perirhinal damage increases object-recognition deficits. Damage to the hippocampus improves performance in object recognition tests, which implies that these two systems are competitive. The hippocampus is also correlated with episodic memory. In human studies the hippocampus may be correlated with whole scene recognition. Rodents have corresponding hippocampal and perirhinal structures, and studies of them mirror those of the primates. The conclusion is that the perirhinal cortex plays a crucial role in object recognition, both perception and memory, that requires complex correlations between features. The hippocampus plays little role in object identification.

This paper furthers the argument for a functional separation of brain regions specialized for object-centric and spatial processing. These representations are not only visual but are multi-modal. A link is made between spatial processing and episodic memory, which makes sense if we consider time as another dimension of space. The recognition of the arrangement of multiple objects in space is correlated with the hippocampus. These features of visual processing are relevant to a developing conceptual system meant to resemble that of an infant.

[37] Y. Nir and G. Tononi. Dreaming and the brain: from phenomenology to neurophysiology. *Trends in cognitive sciences*, 14(2):88–100, 2010.

Nir and Tononi provide an alternative to Hobon's bottom up conception of dreaming. For Nir and Tononi, Hobson's argument regarding the importance of PGO (pontine-geniculate-occipital) waves in dreaming is weak because PGO waves only occur in REM sleep, but studies have shown that dreams also occur in NREM sleep. Nir and Tononi contend that dreaming is not a bottom up process, where random activation stimulates perception, but that perceptual systems are activated by higher level brain functions. The crux of the theory is that dreams are like imagination.

Hobson, Nir and Tononi largely agree on the qualitative features of dreaming, corresponding to Hobson's conception of primary consciousness. A distinction is made between dreams that occur in REM sleep and those that occur in NREM. REM dreams have features described by Hobson, including reduced volition, self-awareness, memory deficit etc.. Reports of dreams occurring in NREM sleep tend to be short, similar to individual thoughts, less visual and vivid, more conceptual, less emotional, more volitional, and more plausible. The frequency of NREM dream reports are rare compared to REM. A key point is that the quality and content of the dreams match the cognitive abilities of the dreamer. For example children dream quite differently than adults and people blind from birth do not dream in visual images. The degree of consciousness in various sleep states is variable. In terms of brain activation (using brain imaging and EEG methods) REM sleep and quiet waking are extremely similar, while NREM is significantly different. There is evidence for consciousness occurring during NREM sleep.

Features of dream content are correlated with mental imagery. In studies, lucid dreaming subjects report a lack of fine detail in dream images. The same lack of detail is reported in studies of mental imagery. The qualities of children's dreams matches their mental imagery abilities. Younger children tend to have less narrative and more event oriented dreams. Very young children seem unable to dream of continuous transformations.

The theory provided by this paper provides a strong alternative to the bottom-up theory proposed by Hobson. Nir and Tononi don't directly address the problem of dream function, but concentrate on the features of dreaming and their cognitive correlates. The theory is relevant to this research program because the top-down conception has significant creative and aesthetic potential. Moving the activation process from low level perception into a high-level conceptual area is consistent with Gabora's conception of human creativity, making a strong cognitively oriented connection between dream experience and creative thinking, which is further reinforced by earlier discussions of hemispherical asymmetry.

[38] A. Oliva and A. Torralba. The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12):520–527, 2007.

This paper discusses the importance of context in object recognition. Elements of the background around an object can provide valuable information, and facilitates more accurate and fast object recognition. Useful aspects of context include semantics, spatial organization and orientation. Contextual cues are most significant during quick glaces of less than 200ms. Temporal context is also important: the recognition of loaf of bread is faster when presented after an image of a kitchen counter, and slower when presented after a bass drum.

Human subjects can learn contextual relations between arbitrarily associated objects. Contextual cues are not impaired by a stretch transformation, but are significantly impaired by changes in viewpoint. What is context? If context is considered object-object relations then perceptual atoms are objects, which require object-centric processing. An alternative is the consideration of context as statistical properties of whole scenes, for example: Mean size and variance of objects, centre of mass, texture, clutter, depth and perspective.

The features of context described in this paper, including arrangement of objects and low-level statistical features, could be useful in Dreaming Machine's perceptual system. The aim of the artwork's perceptual system is not the recognition of an object, but to allow a conceptual system to group a number of perceptual units (based on object and/or spatial features) under an overarching concept, of which recognition is an aftereffect.

[39] J. P. J. Pinel. *Biopsychology*, chapter The Visual System. Pearson Education, 2009.

This chapter provides an overview of the visual system, which exhibits an "inherent creativity". This summary bypasses eye and retinal anatomy, edge enhancement and colour perception to focus on the retina-geniculate-striate and the visual cortex. The retina-geniculate-striate is the pathway that passes signals from the retina, through the lateral geniculate nuclei (LGN) of the thalamus, to the striate—the primary visual cortex (PVC). Signals from both eyes are interleaved and projected into the PVC in both hemispheres. Throughout the retina-geniculate-striate, signals are topologically mapped so that nearby retinal cells are projected onto nearby striate cells. There are at least two parallel communication channels that pass through the retina-geniculate-striate: The "P" channel connects mostly, colour sensitive, cone cells to parvocellular layers of the LGN. These cells are tuned to colour and details of stationary and slow moving objects. The "M" channel connects mostly rod cells, sensitive to luminosity changes and low light, to the magnocellular layers of the LGN. These cells are tuned to movement and not colour or fine-detail. P and M channels project to slightly different parts of the PVC.

Receptive fields are circular monocular regions of the visual field in which a particular neuron's firing can be influenced. Each neuron then has a receptive field and is only sensitive to signals corresponding to that region. The size of the receptive field does not change through the retinal-geniculate-striate, which implies little lateral interaction. There are two classes of receptive fields in the striate: (1) Simple cortical cells have monocular receptive fields and are organized into lines. These cells respond to edge contrast along lines in particular orientations. (2) Complex cortical cells do higher level image processing and are more numerous than the simple cells. These cells are also sensitive to contrasting edges, at certain orientations, but have larger receptive fields and are often binocular (but can also exhibit dominance to one eye). Complex cells respond to edges of particular orientations independently of their location in the visual field. The behaviour of simple and complex cortical cells lead to the theory of columnar organization of the PVC. According to this theory, signals from simple cells are combined in complex cells and are functionally grouped into the columns in the PVC. Cells in a single column are tuned to the same general area of the receptive field, and to edges of a particular orientation. As one examines the receptive fields of neighbouring columns by sweeping along the PVC it is clear that columns are organized both by the location in the receptive field, and the orientation of edges to which they are sensitive. Columns are topologically organized in that nearby columns have similar receptive fields, and are tuned to similar edge orientations. Areas of left and right eye dominance are interlaced through the columnar structure.

Once visual information is processed by the PVC it is distributed to higher levels of the visual system via two pathways: according to Ungerleider and Mishkin the dorsal pathway is sensitive to the position of objects, while the ventral pathway is sensitive to the appearance of objects. Different neurons in the ventral stream are tuned to different classes of objects, such as faces, letters, animals and so on. According to Goodale et al, the dorsal stream integrates sensorimotor systems in the intentional manipulation of objects, while the ventral stream performs the same function as proposed by Ungerleider and Mishkin.

Just as knowledge of neuron anatomy can inspire biologically oriented artificial intelligence methods, knowledge of the visual system can inspire methods in computer vision systems. The system produced during this research is expected to process sensory data from the world in order to create conceptual structures that represent it. This requires a computer vision system, and one inspired by biology is preferred. In particular, the functional separation of location and object based visual processing is highly relevant.

[40] J. P. J. Pinel. *Biopsychology*, chapter Mechanisms of Perception. Pearson Education, 2009.

High level concepts regarding the human perceptual system are discussed in this chapter. This summary only includes key points relevant to this research program—pain and attention. The perceptual system is hierarchical and functionally separated. Signals from sensory neurons are not only abstracted, but high level abstractions can effect sensory neurons. Interactions between sensory systems happen at many different levels of abstraction, not just at high "associative" levels. Pain is a component of the somatosensory system and is not well understood. Pain may exist to assist in survival, forcing the organism to avoid activity during healing, and avoid damage. Pain is not well represented in the cortex—its perception is not tied to any particular cortical location. Pain can be suppressed through cognitive and emotional factors.

There are two types of attention: (1) Endogenous attention allows high level cognitive processes to influence where attention is placed, for example watching something in your visual periphery. (2) Exogenous attention is when an external stimulus forces our attention to a particular event, for example a sudden loud sound forcing our immediate visual attention. Neurons in the visual system may selectively inhibit areas to which the organism is not attending.

Pain may be a used as a primitive emotion that could be used to drive intentional and motivated behaviour in a system. This is important as some models of human development require goal-directed behaviour for the generation of concepts, and therefore self. A neurological understanding of pain would allow a consideration of some machine analogue. Unfortunately, pain is not well enough understood to provide an non-arbitrary mapping, on a neurological

basis, to a machine. Attention may be intrinsic to perception and could therefore be useful in a computer vision system. The possibility that an high level attentional mechanism may alter low-level sensory systems gives a hint as to how imagination could result in mental imagery. High level neural structures may effect low level neural structures as much as low level strictures effect high level structures. This relates to Piaget's differentiated signifiers that become detached from the sensory patterns that initiated them.

[41] J. P. J. Pinel. *Biopsychology*, chapter Neuron Conduction and Synaptic Transmission. Pearson Education, 2009.

This chapter provides an overview of how neurons communicate via electrochemical signals. A neuron has a resting membrane potential, that is a voltage (potential) difference between the inside and outside of the cell. The neuron is polarized while in this resting state where the potential is approximately -70mV. When neurons fire they release neurotransmitters through buttons that are received, commonly, by the postsynaptic cell's dendrite. The receipt of these neurotransmitters can either increase (hyperpolarize) or decrease (depolarize) the postsynaptic cell's potential. Excitatory postsynaptic potentials (EPSPs) is the state where the postsynaptic cell is depolarized, and increases the likelihood that the neuron will fire. Inhibitory postsynaptic potentials (IPSPs) are the opposite, where a hyperpolarization decreases the likelihood that the neuron will fire. The amplitude of these potentials are proportional to the intensity of the triggering signals.

Neurons are often connected to thousands of other neurons. The PSPs from all presynaptic neurons are summed (integrated) and if they exceed the threshold of excitation, the postsynaptic neuron will fire. A PSP from a single presynaptic neuron has little effect on the firing of the postsynaptic cell. When a neuron fires it generates an action potential (AP) not proportional to the sum of presynaptic PSPs. The integration of PSPs occurs in both temporal and spatial terms. Two simultaneous PSPs are summed as are two in close succession. Once a cell is depolarized by a EPSP, ion channels open and positive ions flood in, which sets off a reaction that results in an AP. After AP has occurred, the cell ends up in a slightly hyperpolarized state. A cells "refractory" period is the amount of time required for a fired cell to be able to fire again. During the absolute refractory period no PSPs can trigger an AP. During the relative refractory period the cell can be triggered, but high amplitude PSPs are required. Once both the absolute and relative refractory periods are complete a normal PSP can result in an AP. Interneurons do not conduct APs (firing potentials), but only PSPs (graded potentials).

There are multiple types of synapses grouped into two classes, directed synapses and undirected synapses. In directed synapses one cell sends neurotransmitters that are received by a second cell. Axodendric synapses are the most common type, where the axon of one cell connects to the dendrite of another. Axosomatic synapses are also common, but axons connect directly to cell bodies, and not dendrites. The less studied dendrodenridic synapse connects two dendrites, while the least common axonaxonic synapse connects two axons. Undirected synapses do not direct their neurotransmitters to a single cell but produce a wash of neurotransmitters that are received by many cells. There are two types of neurotransmitter, small and large, and many neurons contain both. Large neurotransmitters are short protein (amino acid) chains called neuropeptides. Exocytosis is the process by which small neurotransmitters are released in bursts and often triggered by the AP. Neuropeptides tend to be released more slowly and in response to the rate at which the cell is firing.

Each neurotransmitter is matched with specific receptor types, of which there are two classes: (1) Ionotropic receptors are rare and change the potential of the receiving cell by opening or closing ion channels. The type of receptor defines whether a particular neurotransmitter should inhibit or excite the postsynaptic neuron, which means that the same neurotransmitter, triggered by a single neuron, can inhibit one cell while exciting another. (2) Metabotropic receptors cause a "G" protein to be released in the cell body that can either cause ion channels to open or close, or can set off a secondary process, for example issuing a change in the DNA of the cell. Cells may have autoreceptors that are matched with the cell's own neurotransmitters. These allow the "reuptake" of unbound neurotransmitters, and may have other functions. Some unbound neurotransmitters are degraded by enzymatic processes. Neurotransmitter processes are not the only way of cells to communicate. "Gap junctions" transport small molecules, and between compatible cells electrical signals, between cell bodies. Unlike neurotransmitter communication, gap junctions are bidirectional and fast. Astrocyctes are known to communicate directly with neurons through gap junctions. They may be integral to local inhibition and their speed may allow temporal syncing of neuron firing.

Over one hundred small neurotransmitters, in three classes, have been identified. Amino-acid neurotransmitters (for example glutamate, GABA, etc.) are fast acting and used in directed synapses. Monoamine neurotransmitters (for example dopamine, epinephrine, etc.) are common in the brainstem and are used in undirected synapses. Acetylcholine is common all over the nervous system, including at neuromulscular junctions, and may be linked with memory processes. There are a number of "unconventional" neurotransmitters: Soluble gas neurotransmitters, such as carbon monoxide, are very short lived and diffuse directly through cell membranes. Endocannabinoids tend to inhibit presynaptic cells. There are over one hundred identified neuropeptides that are found in various locations, including the pituitary gland, the thalamus, and the gut.

This chapter provides detailed information on how neurons (and glia) communicate. This has a bearing on connectionist AI approached inspired by biology. In particular the background in neurotransmitters is valuable in the context of neurotransmitter changes that occur during sleep and dreaming states. The function of neuron communication is so complex as to make any implementation difficult. Nevertheless the overall concepts of neuron communication could be useful in connectionist systems.

[42] J. P. J. Pinel. *Biopsychology*, chapter The Anatomy of the Nervous System. Pearson Education, 2009.

This chapter provides an overview of the anatomy of the nervous system. This summary focuses on neuron anatomy due to its relevance for artificial neural networks. Neurons are cells specialized for the reception, conduction and transmission of electrochemical signals. Neurons are composed of four major components: (1) The body is the metabolic centre of the neuron. (2) Dendrites receive signals from multiple cells. (3) The axon is a long string that terminates with "buttons" that store neurotransmitters in synaptic vesicles. (4) Synapses are the gaps between the receiving cells' dendrites and the sending sells buttons. Neurons can be grouped into four classes: Unipolar neurons have a single projection from the cell body, while bipolar neurons have two projections. The multipolar neuron, the most common class, has multiple projections. Interneurons have multiple dendrites and "integrate" neural activity by relaying signals between neurons. In addition to neurons are glial cells that outnumber neurons in the human brain ten fold. The known functions of glia continue to grow, three of the most known types are: Oligodendrocytes generate the myelin that insulate the axon. Microglia extend immune responses by removing debris, signalling inflammatory responses etc.. Astrocytes interface with blood vessels and neuron cell bodies. These glia release chemical transmitters, have neurotransmitter receptors and conduct signals.

In previous projects, including "Memory Association Machine" and "Dreaming Machine #1 and #2", connectionist AI was explored due to its "bottom-up" approach, its biological inspiration, and the potential for many simple components to generate emergent behaviour. The study of contemporary conceptions of neurons may provide inspiration in the design and architecture produced during this research project.

[43] J. P. J. Pinel. *Biopsychology*, chapter Sleep, Dreaming, and Circadian Rhythms. Pearson Education, 2009.

This chapter describes the study of sleep and possible functional theories of sleep. Rather than a human specific behaviour, sleep is nearly universal across all animal life. Human sleep studies are most associated with Electroencephalogram (EEG) methods that measure the synchrony of neuron activity. There are four sleep stages (1-4) which correspond to the degree of neuron synchrony in the brain. Stage 1 is the most similar, in activity, to the waking brain, characterized by high frequency and low-amplitude waves that indicate a lack of synchrony. Stage 2-4 (aka SWS) are NREM sleep stages and involve increases in the amplitude and decrease in frequency and indicates an increase of synchrony. A typical sleep cycle begins with stage 1 followed by a strong increase in synchrony leading to stage 4. Once stage 4 has been reached the sleeper oscillates between sleep stages in sequence. When reaching stage 1, after the onset of sleep, REM occurs. For each oscillation more time is spent in REM sleep.

There are two major theories for the function of sleep. Sleep is may be recuperative, and serves to give the mind and body rest, or circadian, where sleep is one of many biorhythms that oscillate over time. The recuperative theory states that waking diminishes homeostasis and that sleep is required to replenish it. There is no correlation between the amount of sleep an animal gets and measures of its homeostasis. The circadian theory states that sleeping is the result of an internal timing mechanism and a possible function is to keep the sleeper safe during vulnerable periods of the day. There is some correlation between animal's sleep time and its vulnerability during sleep. The timing of circadian rhythms are trained from environmental cues or "zeitgebers" for example: light level, social activity and eating. Neither of these theories provides a theory of sleeping that resolves all evidence, and the function of sleep is likely a combination of the two.

The chapter discusses a number of possible effects of sleep deprivation. Evidence supports the idea that the ill effects of sleep deprivation may be due to a disruption of circadian rhythms rather than the decrease of sleep time itself. The majority of animal studies on sleep deprivation involve extremely stressful methods of keeping animals from sleeping. The reported effects of sleep deprivation may be attributed to this stress, rather than the lack of sleep. Common antidepressant drugs stop REM sleep entirely and no significant effects have been found in these patients. The sleeping pattern of the sleep deprived shows an increase in the amount of SWS and a decrease in the amount of REM sleep. The lack of SWS sleep causes the major effects of sleepiness. In long-term studies of sleep deprivation only minor effects, such as the slight lack of auditory vigilance and the feeling of sleepiness, were encountered with decreases of sleep time down to approximately 4 hours. No change in scores related to mood, medical or performance tests.

This broad discussion of sleep research provides valuable insight into the major components of sleep. In particular the discussion of circadian entrainment could be used to implement an entrainment method so that the sleep cycle of the machine could reach an equilibrium with the context of installation. The difference in EEG signals at various sleep stages may give insight into the role of neuron synchrony in sleeping that could be implemented in the Dreaming Machine.

[44] M. Shah. Fundamentals of computer vision. *Computer Science Department, University of Central Florida, Orlando, FL. Disponível*, 1997.

Shah provides a summary of foundational methods in computer vision, including algorithms and mathematical basis. Chapter 3, region segmentation, is most relevant to this research program. Segmentation is the process of relating groups of pixels according to their properties. The chapter discusses three major methods of segmentation: Simple segmentation, Connected Components, Seed Segmentation and Region Growing.

Simple segmentation involves the application of thresholds to select areas of an image. Thresholds are determined using an analysis of the image histogram. First the histogram is smoothed to remove any small spikes due to noise. Second the "peakiness" test is used to select significant peaks. The number of gaps between significant peaks is the number of thresholds, whose values split the distance between significant peaks.

The connected components method involves an analysis of the neighbourhood around each pixel. The process is as follows: scan the image from left to right and from top to bottom. For each pixel with a particular value that is unlabelled assign a new label. For all the unlabelled neighbours, with the same value, around this pixel assign the same label. Labelling for this value is complete once all pixel with a that value are labelled. In "seed segmentation" the

"simple segmentation" method is used to determine thresholds. These thresholds are used to segment the image. Once the image has been segmented the segmentation can then be refined using the connected components method.

Region growing is a family of segmentation methods. The ones discussed in this document include "split and merge", phagocyte, and "likelihood test" methods. Split and merge uses a heuristic rule to determine the relation between pixels. The simplest sample rule is to assume pixels with the same value belong to the same region. Using this example heuristic segmentation proceeds as follows: Take the whole image and determine if all pixels have the same value. If they do then the segmentation is complete, if not then break the image into quads (split). If all the pixels in two adjacent quads are the same then merge those two regions. Once no more merging is possible split the non-merged quads again and continue. Segmentation is complete when all the pixels in the merged region have the same value. The phagocyte method depends on a heuristic of the "weakness" of the boundaries of a region. If a region is weak, then its merged with the enclosing region. The weakness of the region is the ratio between the number of weak boundaries and the total number of boundaries. The weakness of a boundary is determined by the sum of the weaknesses of its edges. The weakness of an edge is determined by the absolute difference in grey-levels on either side of the edge. The weakness heuristic tends to over-merge regions. An additional phagocyte heuristic takes into account region perimeter and counteracts the tendency for the weakness heuristic to over-merge. The likelihood ratio test is a probabilistic method for determining if two regions are the same based on their grey level distribution. A probability distribution of grey-levels is calculated for both regions. The regions are merged if they ratio between these probabilities is below a certain threshold.

The segmentation methods discussed are valuable for this research program for object-centric visual processing which requires the segmentation of objects from the background. A combination of seed segmentation and background extraction should be sufficient for the needs of this project.

[45] D. F. Sherry and D. L. Schacter. The evolution of multiple memory systems. *Psychological Review*, 94(4):439–454, 1987.

This paper provides an evolutionary perspective on the role and evolution of memory systems. The purpose of memory is to allow an organism to acquire, retain, and retrieve information. These features allow organisms to recognize the familiar, predict events, return to particular places, and assess the consequences of previous behaviour. "Functional Incompatibility" states that the functions memory is used for are so diverse that a single system of would not be able to enable them. Different memory systems are characterized by different rules of operation. Systems may be independent, or dependent on other systems. The same information can be fed into multiple memory systems, or different information can be associated with different memory systems. There are different processes that handle different kinds of information, but it is not clear how these processes relate to memory. Multiple memory systems occur where different processes use different memory systems that operate according to their own rules, and do not overlap in function.

Memory should be considered within an evolutionary perspective as the development of memory systems is likely tied to the kinds of environments in which the organism is situated. There is strong evidence that memory systems are effected by genetics. Memory is obviously correlated to reproductive success through its effect on avoiding predation (recognizing predators), "animal contests" (where dominance is tested and maintained), cooperation (recognition of social grouping and lineage) and foraging. Different memory systems may evolve for different survival problems.

In primates and humans, a number of memory systems have been proposed. These systems can be grouped into two categories: (1) Memory system I is characterized by gradual and incremental learning and corresponds to procedural, implicit and dispositional memory. (2) Memory system II is characterized by "single-trail" memory construction and corresponds to declarative, explicit and representational memory. Often memory impairments effect one system and not the other. System I is about feature extraction (preserving invariance across episodes), while System II is about representation (the preservation of variance). Memory of up to a few seconds may be enabled by a third system.

This paper makes two significant points relevant to this research: (1) there is likely no single unified memory system that does all that we attribute to memory (and perhaps learning, not discussed in this paper) and (2) the structure of memory systems evolves for particular survival needs. "Dreaming Machine" is expected to generate a network of concepts from a continuous source of visual sensory data. The structure of this network will depend on particular models of memory systems. Additionally, the artwork is an "alien" entity, it is meant to be similar to a human but is in no way identical. As such, the way its perception and memory is constructed may result from its particular purpose and design and may not resemble human memory in any way. Any mapping between human and computational memory is problematic.

[46] C. Siagian and L. Itti. Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE transactions on pattern analysis and machine intelligence*, 29(2):300–312, 2007.

The authors propose a scene classification method that is biologically inspired and includes a mechanism that could enable attention. Three types of scene recognition are discussed: (1) Object based scene recognition involves segregation, grouping and object recognition to recognize scenes based on the objects present and their arrangement. A problem with this approach is knowing which objects are significant to a particular scene. An example method is the Scale Invariant Feature Transform (SIFT) which does only partial recognition of objects and is therefore more flexible. (2) Region-based recognition is similar to object-based recognition except entire segmented image regions are used, rather than objects. This method suffers from similar problems as object-based systems, in particular the difficulty with segmentation. (3) Context-based recognition uses the entire image as a single unit and does not depend on object recognition or segmentation. The entire image is abstracted using statistical methods, such as histograms. As the whole image is used it is not as sensitive to noise as other methods. Techniques such as texture descriptors, histograms, FFT, PCA, and wavelets are used in this class.

A subset of scene recognition systems are biologically plausible and inspired. Humans' ability to classify scenes is extremely quick and accurate compared to computer vision methods. Fast recognition is based on the "gist" of the whole scene, which appears related to spectral components and colour diagnosticity. Gist methods may be complimentary with saliency based methods that involve more detailed analysis of individual regions. Saliency is associated with the dorsal visual stream where differences are emphasized, while gist is associated with the ventral stream, represents the whole scene, and emphasizes similarities between scenes.

The method the authors propose uses gist scene measures. Images are re-sampled at eight different resolution scales and smoothed using a 5x5 Gaussian filter. Each of these scales is fed into three parallel filter channels. The orientation channel uses Gabor filters at multiple orientations, while centre-on and centre-off filters, corresponding to simple cortical cells in the primary visual cortex, are applied to separate colour and intensity channels. A point-wise absolute difference operator is applied to pairs of images, resulting in a single vector representing differences. These difference images are broken into 4x4 grids and the mean pixel value for each regions is calculated, resulting in a 16 element vector. The dimensionality of the results are reduced using PCA, and the output used to train a multilayer perceptron using backpropagation to classify images. Scenes can be classified effectively using this method which is able to handle "... translational, angular, scale and illumination changes."

The initial conception of the perception system was based on a visual system that was biologically plausible. This drive for biological plausibility, at the neuronal level, has since loosened in order to devote more computational resources to higher functions. Efficient, well-established and perhaps non-biologically plausible computer vision methods will serve as the perception system for the artwork. Statistical summations of low-level features is still a useful technique.

[47] D. Y. Teller, M. A. McDonald, K. Preston, S. L. Sebris, and V. Dobson. Assessment of visual acuity in infants and children; The acuity card procedure. *Developmental Medicine & Child Neurology*, 28(6):779–789, 1986.

This paper proposes a method of measuring infant visual acuity. The "acuity card procedure" uses a standard set of cards with varying frequencies of vertical stripes. The average luminosity of these cards is matched by a reference card. The procedure follows from previous methods based on "preferential looking" which predicts that infants tend to look longer a pattern than a solid colour of the same average luminosity. The proposed procedure has the infant or child at a standard distance from two cards placed horizontally. The researcher observes the infant through a hole in the card and check or preferential looking, and other indications the child can see the card stripes. The assistant/parent holding the child is unable to see the cards, and is therefore unbiased. The procedure was developed because previous PL methods required expensive awkward equipment and many children were not testable. The acuity card procedure is faster, more simple and allows a much higher percentage of children to be tested.

In order to develop an artwork that develops a conceptual system in a similar way to an infant, its perceptual system should contain a similar level of acuity. One of the foundations of Mandler's Image-Schema theory is that motion is the basis of conceptual development because infants do not have the visual acuity to recognize objects based on their non-spatial features. Results of the proposed methodology could be used to develop a simple model of infant visual development that could be used in the "Dreaming Machine".

5 Development, Mental Representation and Consciousness

[48] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida. Cognitive developmental robotics: a survey. *Autonomous Mental Development, IEEE Transactions on*, 1(1):12–34, 2009.

This survey covers the current state of Cognitive Developmental Robotics (CDR), an approach to robotics inspired by human development and neurology that emphases the importance of physical embodiment in the formation of cognitive functions and social behaviours. The creation of machines that gain "higher order cognitive functions through learning and development is one of the greatest challenges in trying to make artificial systems more intelligent. . . ". Most implementations are superficial because: (1) There is "little knowledge and few facts" on how higher level cognitive functions arise. (2) The mapping of these processes to designed machines is unclear. (3) Is there sufficient understanding under the view that higher order functions arise from primary functions? One solution to these problems is to construct whole systems that are not reduced to individual components.

Development can proceed in two stages: (1) early individual development as informed largely by neurology, and (2) social development resulting from embodied interactions with other individuals, as informed by cognitive science and developmental psychology. The authors propose a methodology that focuses on two parallel efforts that are expected to mutually support one and other: (A) the construction of computational models of cognitive development, and (B) to offer new data to better understand the developmental process. (A) involves the generation of hypothesis that are verified using computer simulation and physical agents, such as humans, animals and robots. (B) Involves the use of brain imaging techniques, verification using human and animal subjects, and providing a robot that stands in for a human subject where ethics are a barrier to experimentation.

There are three major aspects of development in CDR: (1) the study of the normal development of the fetus and the infant, (2) the construction of artificial systems and (3) the building of a model of cognitive development. Recent studies of fetuses expose a number of physical behaviours thought to be restricted to infants, including sucking, swallowing, hand-face contact, breathing, yawning, etc. The authors highlight a number of issues in development: (1) Infants brain structure and function cannot be derived by adults, (2) brain regions for development and maintenance are not the same, (3) attention develops from bottom-up processes, such as saliency maps, to top-down processes, and (4) even if performances look similar, the corresponding neural structures may be different.

In terms of artificial systems, development is "...not centrally controlled but instead a distributed and self-organized process" where higher levels emerge from lower levels before they are complete or even efficient. A crucial aspect is that the active exploration and interaction with the environment is regarded as central in perceptual categorization and concept formation. The caregiver provides a "scaffolding" for cognitive functions, social abilities and skill development. The authors propose a model of cognitive development where the brain-spine is considered a hierarchical system that reflects the "evolutionary" process involving the "spine, brain stem, diencephalon, cerebellum, limbic system, basal ganglia, and neocortex". This structure informs the major developmental components: "reflex, sensorimotor mapping, perception, voluntary motion, and higher order cognition"

The motor aspects of development are dominated by the spine, the medulla, the parietal lobe and the cerebrum. These areas contribute to the development of a "body schema". Recent studies show fetuses exhibiting behaviours which are thought to require body-schema. Kuniyoshi and Sangawa have constructed a simulation of fetus development that exhibits a number of these behaviours, which indicate that the fetus brain and body interact before birth. Computational models of early movement and sensor integration, most of which do not involve embodiment, have been proposed by Chen, Izhikevich and Edelman, Kinjo and Pitti.

Similar models that focus on the movement of young infants are proposed by Kuniyoshi and Sangawa, Righetti and Ijspeert. A number of efforts have been made to embed computational models in dynamic and embodied robot bodies: Degallier, Blickhan, Raibert, Hyon, Iida et al, Hurst et al, Vanderborght et al, Hosoda and Ueda, Narioka and Hosoda, Niiyama and Kuniyoshi. The authors propose an integrated biomemetic body for developmental studies called CB^2, which is "...designed, especially to establish and maintain a long-term social interaction between human and robot." The system features a whole body skin embedded with tactile sensors, flexible joints and 51 pneumatic actuators. The flexibility in the joints requires a new control system, designed by Ikemoto, that capitalizes on this flexibility in order to provide a tight physical interaction with humans.

Conventionally, the body representations used in robotics are determined from the parameters of skeletal morphology, that is the arrangement of joints in space. An adaptive method involves the use of environmental experience to allow the system to learn a body representation that is flexible. The flexibility of body representations is well studied by Ramachandran and Iriki. Body representation through multi-modal integration has been studied via robots by Nabeshima, Yoshikawa, Fuke, Asada, Yoshikawa, Stoytchev, Hersch and Hikita.

The role of the caregiver is central in an infant's development of social behaviour and communication. Not enough is known about this process "...when applying the theories to the design of communicative robots", which is an opportunity for robotics to validate theories. The timing of care-giver behaviours appears of particular importance, as explored by Mirza and Ogino who built robots that play peekaboo. Vocal imitation requires the infant to have a sensorimotor mapping to vocalize, and also find the correspondence between their and the caregiver's vocalizations. Computational models in this area are developed by Guenther, Westermann and Miranda, Kanda, and Miura.

Joint attention, the ability for the caregiver and infant to attend to the same object, is a central basis of the development of communication. The infants ability to show preference to faces and face-like patterns is believed to be innate, and yet subjects with autism show no or little preference to faces. Particular behaviours enabling joint attention include gaze alternation, social referencing and pointing. Statistical mapping has been used by Yoshikawa to construct models that combine gaze-following and contingency (using an arbitrary salient object that intersects with the caregivers gaze to constrain search) in combination with word-to-object mapping. Empathy is considered "indispensable for communication". Research shows that as much as fifty-five percent of communication channels focus on the face, while thirty-eight are tuned to the tone of voice, and only seven to verbal content. An important mechanism in this area is the caregiver's imitation of the infant's vocalizations.

The "verbal explosion" is the rapid increase of infant vocabulary from one to two thousand words between approximately 18 and 24 months. Language acquisition is normally approached using a "bottom-up" method where the "symbol grounding problem" (the problem of how to sync sense information from the environment to auditory data from the caregiver) comes to bare. In most of these systems learning is passive, which is not supported by the developmental literature. For example, the onset of the verbal explosion coincides with the infants ability to walk, implying a tie between active embodiment and language acquisition. Curiosity may be an important drive to lead the infant to objects that have not yet been associated with sounds. The newness of objects, in respect to the infants experience, is also a significant drive.

CDR focuses on a tightly coupled and active physical interaction with the environment. This emphasis on the active manipulation of objects overlaps with Piaget. The artwork proposed in this research is not expected or intended to be mobile. This is problematic if the work is expected to rigorously follow Piaget. A central philosophical and technical question answered by this work is whether a conceptual system can form in a immobile body that can only attend to various aspects of the visual field without actively manipulating them. Additionally, the viewer is not expected to assume the role of the "caregiver" of the artwork, which is not intended to exhibit socially interactive behaviours. The system may include a face-detector, which could be innate, and be used as an emotional signal in learning.

[49] E. Thompson F. J. Varela and E. Rosch. *The Embodied Mind: Cognitive Science and Human Experience*, chapter Symbols: The Cognitivist Hypothesis. MIT Press, 1991.

This text describes the origins of cognitive science, defines the "cognitivist hypothesis" and surveys the relation between cognitive science and artificial intelligence, the brain, psychology and psychoanalysis. The chapter concludes with a discussion of the relation between knowledge in cognitive science and experiential knowledge, and discusses Jackendoff's proposal to integrate phenomenological research into cognitive science.

The authors believe that many of the current debates in cognitive science were introduced between 1943 and 53, which indicates their difficulty. The origin of cognitive science is located in cybernetics, a multi-disciplinary effort to study life and minds using technological and mathematical tools. Cybernetics is characterized as "...complex and entangled but also rich with possibilities for growth and development." This "exploratory stage" is contrasted with the contemporary state of cognitive science. Cybernetics made the following contributions through an "intertwining of the philosophical, the empirical, and the mathematical": (1) The use of math to understand the nervous system. (2) The invention of information processing machines that pave the way to AI. (3) The establishment of "systems theory", and (4) information theory. (5) The first examples of self-organizing systems.

Cognitive science became its own discipline in 1956. The central hypothesis is that intelligence in general, human intelligence included, "so resembles computation in its essential characteristics that cognition can actually be defined as computations of symbolic representations." Therefore the notion of "representation" is at the core of cognitive science. Cognition is "...acting on the basis of representations that are physically realized in the form of a symbolic code in the brain or a machine." The problem is then to map representational states with the physical changes the brain undergoes. The hypothesis requires that the mind is a symbolic computer, where symbols are both physically realized and have semantics. In a computer system the semantics are only accessible by humans reading those symbols, or by reflecting their meaning in the syntax of the program. Cognitive science has then introduced a third level of description for the mind-brain: added to the physical and neurobiological is now the symbolic. The central problem is that it is not known how these physical symbols would get their meaning, as they don't have external readers or programmers to ascribe it, which often leads to homuncular descriptions.

AI is a clear continuation of the cognitivist hypothesis, the construction of symbolic systems meant to exhibit aspects of human behaviour, in particular intelligence. In neurobiology, information processing conceptions are prevalent. The philosophical roots of which are rarely considered or questioned. In psychology, behaviourism was dominant for some time, which considered human minds as black boxes which were impossible to penetrate. The effects of mental imagery on perception could not be explained in behaviourist frameworks and cognitivism took over. Psychological theories could now be formulated, and even validated, through computational implementation. According to "strict cognitivism", mental images are still a problem and considered "epiphenomena" by Pylyshyn. Freud believed that nothing could effect behaviour unless it was symbolically mediated. The unconscious is composed of representations that are not accessible to consciousness. This differs from contemporary cognitivism because, for Freud, all representations have the potential to become conscious.

According to cognitivism, there are mental processes (cognitive processes) that we are not, and cannot be, aware of. This leads to the fundamental fragmentation of the self, as certain aspects of self are not accessible. Further, it has been states that if these unconscious mental processes were accessible to consciousness, then they may cease to function. In first person experience, it seems clear that consciousness and cognition are not independent but "in the same domain". Cognitivism does not require consciousness of its subjects, a symbolic computational system that appears to function in the domain it is meant to is sufficient. If cognition can exist without consciousness, then what is consciousness?

Jackendoff characterizes the relation between consciousness and cognition as two disconnected minds. Causally connected to the world is the "computational mind" (cognition), which "projects" onto the "phenomenological mind" (consciousness). If cognition gains meaning through a representational relation to the world, then how can consciousness, as distinct from cognition, access those representations in a meaningful relation to the world? In other words, how can symbols gain semantic meaning if they are disconnected from experience? According to this conception, all experiential states must have correlates in the computational mind. This requires phenomenology to bridge the gap—a "... disciplined and open-ended approach from the side of experience" is required to fill the holes in cognitive science. It follows quite clearly from Jackendoff's split that the phenomenological mind must be causally disconnected from the computational mind. This begs two major questions, what is free-will if consciousness is causally disconnected, and what is the function of consciousness if it has no causal value? The authors conclude that the complete dismissal of experience (cognitive science's lack of requiring consciousness) leads science of mind to an impasse, as also does an unquestioned acceptance of it (Jackendoff's position).

During the writing of the thesis describing "Memory Association Machine" (MAM) the question of meaning came about. It is this question that appears to be the core philosophical aspect of this research. The argument posed by Varela et al. appears quite clear, and explains the odd theory that consciousness may be "an illusion", as it is factored out by unconscious cognitive processes that are purely symbolic. In MAM, a causal conception of meaning was accepted, the simple causal chain from world to representation. This is in opposition to the Piagetian conception of representation: a representation can only be formed by a conscious agent, that actively makes the link between the world and the symbol. In relation to this research, neither of these conceptions is ideal. The first is fully materialist, in physics what kind of meaning could there be that is not causal? The second is phenomenologically consistent, but tells us little about what the "agent" is, and makes consciousness a requirement of machine meaning. Agre's reconsideration of AI, as a methodology of uncovering conflicts and philosophical weaknesses in underlying theory, comes to the forefront. On one hand is the materialist (causal) conception of meaning, and on the other the intentional agent. It is unclear how to proceed, as the chasm between them appears immense.

[50] M. Guhe, A. Smaill, and A. Pease. Towards a cognitive model of conceptual blending. In *Proceedings of the 10th International Conference on Cognitive Modeling*, pages 293–304, 2010.

Conceptual blending, as considered by Goguen, is the construction of concepts through the combination of existing ones. Most accounts of scientific creativity emphasize metaphor and analogy which, according to Fauconnier and Turner, involve two aspects: (1) Links are established between domains of knowledge representation and (2) these links allow the transfer of knowledge between domains. Metaphor and analogy are particular cases of conceptual blending. In the general case, the Linking and transfer of knowledge is extended to include "reconceptualizations" that allow flexible mappings that cannot be accounted for by metaphor or analogy alone, as they remain "purely descriptive". An example of conceptual blending is the concept of the houseboat, which results from the blending of the house and boat concepts linked by a "base domain". Sortal frames limit the number of possible blends between concepts. The authors discuss two options for implementing sortal frames in ACT-R in order to enable conceptual blending: (1) Use production rules to do sortal checks, and (2) change the ACT-R architecture by extending the declarative module to include built-in sortal checking. The choice is to be determined based on empirical data.

Conceptual blending is highly related to conceptual development. Computational notions of conceptual blending can give clues to how the artwork produced during this research could generate its own concepts from sense-data. The notion of conceptual development in this paper is entirely symbolic. It is not clear how such concepts relate to underlying sub-symbolic aspects.

[51] J. M. Mandler. Thought before language. *Trends in cognitive sciences*, 8(11):508–513, 2004.

This paper describes a possible process by which concepts are formed during infancy and provides an argument as to the requirement of a conceptual system to be in place before language can develop. At the core of the argument is a distinction between procedural and declarative knowledge. Procedural knowledge is implicit, does not require concepts, and is inaccessible to experiment. Declarative knowledge is explicit, requires a conceptual system, and is accessible to experiment. The argument is based on experimental evidence of infant behaviour. At the core of the methodology is "preferential looking" where an infant looks at a particular object for a sustained period of time. Preferential looking is assumed to indicate the creation of a perceptual category, not a concept. The creation of perceptual categories is automatic and does not require conceptual thought. "Perceptual Meaning Analysis" is a conscious process that extracts features from perceptual categories and "redescribes" them into image-schemas. These schemas as the basis of language development, and represent the behaviour, rather than the appearance, of objects.

In the "delayed imitation test", measures the ability of infants to recall and replay actions after a delay. Infants 8-9 months can replay an action, cued by an object, as much as 24 hours after initial presentation. Mandler argues this indicates that the infants have the root of a declarative system, as the routine is remembered and can be repeated even after a single exposure and without practice, which is believed to be required for implicit knowledge development.

Preverbal concepts are concepts composed of image-schema components that are not yet associated with linguistic qualities. Prevebal concepts are initially applied very widely and are refined and divided through development. Dishabituation tests have shown that infants of 7 months can differentiate between furniture, animals and vehicles. 9 month olds can distinguish between images of planes and birds (even with the same colour features), but not between images of dogs and fish. Mandler concludes that these results show that, for some visual discrimination tasks, a developed conceptual system is required.

An issue with this paper is the conflation of representation and object. In all the studies cited, images of furniture, animals and vehicles are presented to the infant, not the referents. If the basis of image-schema is indeed the behaviour/movement of objects, then its unlikely that the infant would link the image of a dog and the dog itself, one clearly displays animate movement, while the other does not. Nevertheless, Mandler provides a highly detailed theory of how concepts could be formed that could be implemented computationally.

[52] J. M. Mandler. On the birth and growth of concepts. *Philosophical Psychology*, 21(2):207–230, 2008.

This paper describes how early concepts are formed through development from infancy. Concept formation is thought to develop from broad, possibly innate, categories and are decomposed and refined over time. This refinement leads to a hierarchically organized set of concepts. Each concept does not include the visual properties of what it represents but only contains reference to visual properties (precept). The features used to generate concepts are locations and movements in space. This is argument partially rests on the fact that young infants do not have the visual acuity to recognize objects based on their non-spatial features. The meaning of objects comes from how they behave in space, rather than how they look. "Perceptual Meaning Analysis" is the process by which perceptual spatial features (Image-Schema) become "redescribed" into concepts. This process requires attention, whereas perception does not. According to the theory, a network of concepts is required for conscious thought. The early concepts constructed during infancy are extended into non-spatial realms and serve as the basis of all concepts formed through adulthood.

According to Nir and Tononi the development of dreams in children correlates with the development of mental imagery, and not language or memory. A correlation with mental imagery suggests that dreams are constructed from a developing conceptual system. As such, theories regarding how conceptual systems are formed, through development, are central to this research.

[53] J. Marshall, D. Blank, and L. Meeden. An emergent framework for self-motivation in developmental robotics. In *Proceedings of the 3rd international conference on development and learning (ICDL 2004), Salk Institute, San Diego*, pages 104–111, 2004.

Marshall provides a framework for self-determined motivation that is based on two "co-evolutionary" drives: the need to accurately predict the environment and the need to constantly seek out novelty in the world that conflicts with learned prediction. This framework is situated in the context of developmental robotics that, as described by Weng, are control systems that are task-independent. Reinforcement learning is often used because it provides feedback without specifying how the problem should be solved. It is unclear if the reinforcement signals should be internal or external. One possible internal source is a model of emotion that would allow states such as pain and restlessness to sequence innate behaviours. The method proposed aims to create a general purpose learning system with self-motivation at its core. The conflicting desires for novelty and predictability are expected to result in emergent complexity and new behaviours.

The method is based on an ANN, a Simple Recurrent Network, and learns from the differences between expected and actual results of behaviour. "Complimentary Reinforcement Backpropagation" combines backpropagation and reinforcement learning. Positive reinforcement is manifest when a stochastic transformation of output units is converted to binary. Negative reinforcement is the compliment of that same stochastic transformation. The amount of noise included in the transformation is the result of a self-regulating "temperature" control. The ANN is composed of output units for motor control and sensory prediction, and input units for motor feedback, sensory and context input. Input and output units are connected through a single hidden layer. In experiment, the system was tested as a virtual robot, with only angle and speed control, in a virtual environment containing a static "decoy" and a moving target object. The goal is for the robot to track the moving object. In the first run, the reinforcement signal was calculated from the sensory representation and stochastic values were suppressed. The robot was able to accurately track the moving object in few training cycles. In the second run, the reinforcement signal was calculated from the robots prediction, rather than the sensory input. The author's found that the robot would only track the object, and learn to do so much more slowly, when learning with the stochastic method. The system achieves a peak performance that was reasonable, but that performance decreased over successive training cycles.

The authors provide a novel method for self-generated motivation. Motivation is considered an emergent result of two underlying and conflicting drives: the need to seek out novelty and the need for stable predictions of the world. This notion is highly relevant to this research because of the importance of motivation in Piagetian conceptions of human development. A system that self-generates motivation solves the problem of implementing predetermined drives. Unfortunately the implementation discussed is only passably successful—considering the extremely restricted test environment. It is unclear if prediction oriented reinforcement contributes anything more than stochastic variation. Implementation aside, the consideration of motivation as resulting from the tension between novelty and predictability could be useful for this project.

[54] J. T. Monterege. The creation of digital consciousness. *SIGART Bull.*, 109:30–33, July 1989.

The appearance of consciousness, rather than restricted intelligence, is considered the most important quality of an intelligent machine. Monterege believes that a machine could manifest consciousness through behaviour that communicates an evolving knowledge about the systems environment, its users and itself.

This paper returns to the notion of the "real goal" of AI, the creation of an intelligent machine, due to two major methodological issues: (1) Mistaking a test for the property it measures. Even if a system passes a test, that does not mean that the system has the property being measured—systems can be built to pass tests that don't exhibit the intended functionality. (2) The transference of a measure of a property in one system to another. Even if there is a correlation between the number of nails a home-builder can hammer per minute and his/her proficiency, it does not mean that any robot that hammers just as well is a proficient home-builder. The functionality of a system can only be determined through careful observation of the system, and its impact on its ended environment. This observation must occur throughout the entire life-cycle of the system.

Consciousness is equated with intelligence—a successful system that behaves in collaboration with humans, for an extended period of time, will appear to the users as conscious. Consciousness is defined as adaptive knowledge in five areas: (1) Knowledge about the world, at a primary school level of understanding, and through elementary level language. (2) The human as a class. (3) The users, as instances of the class of humans. (4) The computer system as a class. (5) Itself, as a instance of the class of computer systems. The first four areas are adapted as a result of user input, communicating information about the world, and reactions to the systems outputs. Self-knowledge is adapted through the system's analysis of its own outputs. The size of the initial knowledge base need not be so large, as long as the system exhibits learning and adaptability.

The system described above is a scaffold for a communicating machine. In order to communicate intelligently it must support a level of communication sufficient to transmit the world and human knowledge. The function of such a system is to be "... an adaptive control system whose function is to optimize, through written communication, behavior in its users that denotes belief and acceptance that they are dealing with an intelligent entity." In short, its purpose is to optimize the user's belief that it is a conscious and intelligent system. Monterege believes that the goal of creating a conscious machine is attainable using the existing technology of the time and is enabled by concerted research efforts in knowledge representation, communication and behaviour.

The conception of consciousness in this paper is deeply rooted in a symbolic AI tradition which is not highly compatible with this research project. The importance of the long-term and real-world deployment of systems is highly relevant as this work is intended for long term public exhibition. The relation the viewer forms with the work is meant to be built slowly over a long period, just as the Monterege's intelligent machine will exhibit consciousness through long term interactions with users.

[55] U. Müller and W. F. Overton. How to grow a baby: A reevaluation of image-schema and Piagetian action approaches to representation. *Human Development*, 41(2):71–111, 2000.

This paper provides a detailed Piagetian account of how mental representation could arise in development. The authors provide a critical discussion of Mandler's image-schema theory in order to argue for their reconsideration of the Piagetian account. The author's emphasize the transition between preoperational (first and second order practical operations) and operational (first and second order representational operations) stages. The central position of image-schema theory is that the mechanisms that allow mental representation are innate. This conflicts with the Piagetian theory, where mental representation grows incrementally through embodied and intentional action in the world. Mental representation is the "action of deploying symbols and signs as referents to absent objects or events". A thinking process is the conscious manipulation of these representations.

For Mandler, the major issue with Paiget's account is the process by which sensorimotor schemes (implicit knowledge) are transformed into representational (explicit) knowledge: how can a conscious process use inaccessible implicit knowledge to generate explicit representations? For Mandler the onset of representational thinking occurs very early (3-4 months). Image-schemas are consciously constructed, by a process of "perceptual analysis", from perceptual information and are the basis of explicit representations. Image-schemas are simplified redescriptions of spatial information and are not directly accessible by consciousness. They form the link between perception and symbolic / linguistic systems. The authors equate image-schema with mentalese and state that it fails to account for the meaning of representations.

Mandler's position comes from an evaluation of Piagetian theory. She argues that there is a Boolean relation between implicit and explicit knowledge. The authors counter this argument by describing the Piagetian process by which signifiers become more and more differentiated over time. This gradual change is mirrored in the increase of consciousness of differentiated signifiers. In Piagetian theory, there is a smooth and gradual transition between implicit and explicit knowledge.

There are a number of conceptual issues with Mandler's theory. The authors find fault in the conscious process that generates image-schemas because perceptions change depending on the task. Mandler provides no evidence that infants experience the process of perceptual analysis, nor that it is influenced by goal-directed behaviour. The conception of the "image" as a carrier of abstract conceptual information is problematic. The theory conflates properties of perceptual objects with properties of conceptual representations. The dependence of the theory on non-intentional perception means that all meaning is causally defined in the world, without the agent's interaction.

The paper provides an alternative constructivist solution that is consistent with Piaget. The theory breaks the transition between preoperational and operational stages into four steps. First order practical operations indicate infant's ability to differentiate on a single dimension. Second order practical operations allow differentiation over multiple dimensions, and are composed of sets of first order operations. At this stage (approximately two years) differentiation is progressed to the point that actions can be executed without the cue. This is because the actions have become broad and can be applied to multiple objects. The next two steps are at the start of the operational stage. First order representational operations occur when the child starts to map between features of multiple concepts. This happens consciously, and requires fully detached signifiers. Second order representational operations involve more complex relations between concepts, including one to many and hierarchical relations.

The evaluation of Mandler provided is largely rooted in the differences between empiricist and constructivist theory. Further, the authors' interpretation of the "image" in an image-schema is incorrect, as an image-schema is not an image but a collection of movement / behavioural properties of an object. As image-schema does not require intentionally directed behaviour, in the construction of a representational system, it is more appropriate for this project than a Piagetian approach. The Piagetian notion that a representational system grows incrementally could provide significant artistic possibilities.

[56] U. Müller, B. Sokol, and W. F. Overton. Reframing a constructivist model of the development of mental representation: The role of higher-order operations. *Developmental Review*, 18:155–201, 1998.

This paper provides a summary of Piaget's account of the development of mental representation. Additionally, it provides a slight reconsideration of Piaget's ideas at the transition between preoperational and operational stages through an introduction of first and second order operations. This paper takes a constructivist position where mental representation is the result of the integration of sensorimotor schemes that are intentionally directed toward the world. In constructivist theory infants have little innate structure and much of mind develops through embodied and intentional interaction with the world. A mental representation is made up of three components, the "representational content" (the signified), the "medium" (the signifier) and the connection (the connection between medium and content). The connection between content and medium results from the agents intentional action—it's is not causal or arbitrary. An alternative to the constructivism is empiricism, where mental representation may be innate, result from perception (rather than sensorimotor integration), and an agent is not required to form a link between the content and medium. The premise that minds and computers are similar (as they are both "physical symbol systems"), is not supported by a constructivist position. The authors argue that computational representations are information and have no meaning as representations are not tied by the intentional actions of the agent to their content. If the representation can be treated as independent of the content then the signifier has lost its meaning, which is derived from the content. Complex syntax does not necessarily lead to semantics.

Piaget's conception of the development of mental representation is divided into two major stages. Before the age of two, infants are in the preoperational stage, where they have not developed mental representation. Once mental representation has developed the infant has reached the operational stage. Assimilation is the process by which the infant transforms the environment to match his/her internal structure: Every object is an object to be sucked. Accommodation is a complimentary process through which his/her structure is transformed to match the structure of the world. During the preoperational stage, signifiers can only be active if the signified is perceptually present. In the operational stage, signifiers are "differentiated" by assimilation and accommodation, where they can be active, even if the signified is not present. This is the genesis of representation, required for reasoning, mental imagery, and thought. Deferred imitation facilitates signifiers to be further and further differentiated as they become used in more and more contexts. The authors argue that this is problematic, as it introduces a paradox: differentiated signifiers are generated by deferred imitation, which requires differentiated signifiers.

The authors resolve this paradox by introducing additional developmental steps between preoperative and operative stages. First order operations begin around the second year of life when infants begin to group objects of a certain class. An example of second order operations is when, later in the second year, children start grouping multiple classes of multiple objects. Grouping in first order operations implies knowledge of similarity between objects along one dimension, while second order operations require grouping along multiple dimensions. Second order operations are sets of first order operations. The authors argue that this is not yet representation and still belongs in the preoperational stage. The notion of object permanence results from notions of displacement and location, also required for second order grouping. The notion of an object itself, independent of its visual features, is an extension of object permanence. During the increasing detachment between objects and actions, actions become the routines that constitute pretend play, where an action is applied to an object with which it is not normally associated. Differentiated signifiers result from pretend play, where actions can be applied to a broader and broader pool of objects. Rather than the object causing a stimulus-response reaction in the child, the child begins to initiate actions independent of objects. Pretend play increases in complexity, as does the knowledge system of the child, as sequences of actions are applied. During second order operations, sequences of actions are planned, which require that operations are considered along multiple dimensions. The complete differentiation of signifiers, when any action can be applied to any object, paves the way for language development. An infant's use of a word to describe an object is the result of first order operations. First and second order operations are preoperational and do not require mental representation. Representation does not arise until the infant is able to explicitly grasp the relation between the representation and the represented.

Piagetian theory is a highly valuable account of the development of mental representation. The authors provide a slight reconsideration of Piaget's theory, but agree on most major points. The constructivist position is relevant to this work because of the lack of emphasis on innate mechanisms and the importance of embodiment are congruent with this research project. That being said, the conception of embodiment present in the constructivist position is difficult to resolve with computational implementation: innate biological needs drive intentional behaviour. The computational analogues of these drives are unclear. The central importance of sensorimotor integration requires a highly embodied system, which is outside of the scope of this project. The final difficulty is the central generator for mental representation, intentional action in the world, requires both biological drives and embodiment. The aim of this research project is not the construction of a complete model of infant development, and therefore these requirements are impractical. An approach aligned with Mandler, which the authors place on the empiricist side, may be a more appropriate account of mental representation for this project.

[57] J. Piaget, B. Inhelder, and P. A. Chilton. *Mental Imagery in the Child: In Collaboration with M. Bovet (o. fl. a.).* Routledge & Kegan Paul, 1971.

Piaget et al. provide a deep account of the characteristics of mental imagery in children and their role in development. Due to the scope of the work, this summary will focus on the general conclusion and not delve into methodologies or arguments. The authors make an early distinction between "reproductive" and "anticipatory" mental images (hereafter referred to simply as images). The former are images that have been perceived by the subject, while the latter are constructions that have not. The development of mental imagery appears to coincide with the transition between preoperational and operational stages. In preoperational children, under 7-8 years, images are only reproductive and static, while operational children are able to imagine kinetic and transformational anticipatory images.

Images have a few interesting limitations: (1) It appears that it is impossible to visualize lines without thickness or points without area. (2) Continuous motion can not be visualized except in a limited fashion as a set of nearly static images that reduce the movement. The reason for (1) is that images must have some correspondence to perception: zero dimensional points cannot be imagined, because they are imperceptible. The reason for (2) is not rooted in perception, as we can perceive continuous movement, but results from a process independent of perception—accommodation.

Images cannot be reduced to perceptions because the boundary effects in images (pseudo-conservatism) do not match the boundary effects in perception. Images must then be constructed through accommodation, where the complexity of images correspond to the growing conceptual structure of the child, which explains the simplicity of preoperational images. The pseudo-conservatism discussed above is explained by interactions between the conceptual structure and the image. In preoperational children perception rules over thought (conceptual structure). It is only through operations that children have the facility (a sufficiently complex conceptual structure) to transform reproductive images into kinetic anticipatory images.

Images do not contribute to the development of the operations, as they are rooted in the static preoperational mode of thought. They may even retard the development of operations as images have the capacity to "...modify the perceptions that serve as their models...". Only once an image becomes anticipatory does it contribute to the operations as they promote the comprehension of transformations.

Language alone does not fulfil the requirements of the semiotic function of thought as it is socially oriented and cannot express all concepts efficiently. Additionally, past events are recalled through imagined images, not through linguistic representations. The authors describe three arguments for the symbolic function of images: (1) The degree of resemblance between images and perceptions varies largely and results from the schematization process (the transformation of images to match conceptual characteristics). (2) While the function of the concept is to interpret and comprehend, the function of the image is to designate. When an image is schematized, changes are subordinate to that which has already been conceptualized. Concepts, words and images are independent, but interacting, systems that compliment one and other. Concepts are represented by verbal signs, while objects are represented by images.

The authors support a knowledge-as-assimilation over knowledge-as-copy position. If images result from perception, then they are copies of perceptions, rather than the objects to which those perceptions refer. The resulting images then have no epistemological value.

Additionally copies are only valid at particular scales and therefore the process "...fails to reach the object...". If an image is broken down into components, then those images require schematization. If we follow the knowledge-as-copy position, then thought is simply a causal result of perception for the purpose of expressing or preserving perceptions. Knowledge-as-assimilation, on the other hand, proposes that "...knowing an object means acting upon it in order to transform it, and discovering its properties through its transformations." Knowledge is constructed through an interplay between the developing conceptual system and perceptions of the world. Both the impact of structure on the world, and the worlds impact on our perception, are interdependent in regards to knowledge forming. The conceptual structure (the "logico-mathematical framework") is required for both the interpretation of perceptions and the construction and evocation of images, which are "indispensable" to the "dynamism of thought".

The account of the development and properties of mental images provided in this book is of significant value to this research. Not only does it fit mental images into Piaget's constructivist approach to development, but also provides empirically derived features of mental images that could be used in the artwork. The possibility that conceptual systems could have their origination in the movement of objects, according to Mandler, contrasts with the static nature of mental images, according to Piaget. Movement oriented concepts may result in the shift from static preoperational images to kinetic operational images, but the time gap (5-6 years) between early concepts, according to Mandler, and the onset of operational thought is problematic. The independence of linguistic and image-based symbolic systems could allow "Dreaming Machine" to be a non-linguistic purely visual "thinker".

6 Artificial Intelligence and Artificial Agents

[58] P. E. Agre. The soul gained and lost: artificial intelligence as a philosophical project. *Stanford Humanities Review*, 4:1–19, 1995.

Agre presents a detailed discussion of the philosophical foundations of AI and proposes a methodological broadening that involves the use of critical and reflexive practises from the humanities. Agre characterizes many AI practitioners as not holding the humanities in high regard. A distinction is often made between "doing", proving theorems and writing software, and "just talking", long term thinking and discourse. AI is considered a self-contained technical field that involves the practical solving of well defined problems. The day to day practise of AI is the formalization, elaboration and implementation of ideas, not reflecting on the value or philosophical background of those ideas. The latter is left to other disciplines that do not have abilities in the former and can do little more than "gesture in the general direction of an idea". Philosophy of mind is discussed amongst AI specialists, but this work is not considered central to AI.

AI systems are often implemented as rules defined by mathematical formalisms. Dreyfus criticizes the use of rules for practical activities when they are applied independently of the culture within which these activities take place: "[t]he attempt to fill in the missing 'background knowledge' through additional rules would suffer the same problem and thus introduce a fatal regress." This kind of critical reflection "is largely a prerogative of the field's [AI's] most senior members". AI's emphasis on formalization and implementation constitutes a powerful method of enquiry because they can force the internal "tensions" to the surface through prolonged technical "frustrations". The analysis of these frustrations, which provide clues as to "...the nature and consequences of the philosophical tensions that generate them", must proceed outside of the technical boundaries of AI and yet can provide valuable insight to future AI developments.

Agre traces the history of Cartesian dualism through to current conceptions in AI. For Descartes, machines (and animals) can never achieve the intelligence and flexibility that constitute the human "soul", or mind. A machine may be able to speak, and perhaps even respond to touch, but it could not "reply appropriately to everything that may be said in its presence".

A machine could only respond in particular and contrived circumstances, never in a general way. While animal and machine behaviour can be reduced to physical laws, human behaviour cannot. Lashley argues against behaviourist psychology's ability to explain complex human behaviour and applied linguistic conceptions of syntax to human behaviour. Sentences can be modelled as sets of rules for grouping words of various classes to generate an infinite number of sentences from a finite vocabulary. The rules correspond to the "bodily capacities" while the generation of sentences result from the mind. As long as AI continues this tradition of dualism, then systems will always depend on both a body and a "soul".

Newell and Simon wrote a program to make choices in a generative space—a set of rules for combining atoms of activity. The soul is considered an epiphenomenon and "... was contained by the generative structure of the search space and manifested through the operation of search mechanisms." Intelligence is simply a search in the generative space. The results of the search need not be perfect, but simply correct enough to allow a search to eventually reach an optimal answer. The problem with this notion is that as the complexity of the world increases, so does the size of the search space at each step, leading to ever increasing search times between steps. Fikes and Nilsson created STRIPS, a program that uses a model of the world implemented as a set of rules in predicate logic, and plans a path through a maze of rooms. Each action is bodily in nature, and the sequencing of these actions results from the will, or soul: "intelligent action emerges from a mass of readily mechanizable decisions."

The above two paragraphs summarize "a single strand of intellectual history", from philosophy and through to technical implementation. The search conception of intelligence is not a failure, but exposes the central problem of dualism—"the soul's causal distance from the world of practical action." Yet this conception of intelligence is still an active area of research, as the philosophical difficulty is not apparent to those looking for technical solutions. Humanities have the potential to expose this philosophical problem, and could provide valuable methods in AI. The strict emphasis on mathematical formalisms is one of the barriers inhibiting integration of humanities because "formalization becomes a highly organized form of social forgetting—and not only of the semantics of words but of their historicity as well." If humanities and AI could be integrated then this formalization would be balanced with a critical eye for the metaphors used to create them. AI could then be characterized by an "... awareness of the cycle of formalization, technical working-out, the emergence of technical impasses, the critical work of diagnosing the impasse as reflecting either a superficial or a profound difficulty with the underlying conception of action, and the initiation of new and more informed rounds of formal modeling." This reformulation of AI would employ the tools of critical enquiry to "engage in a richer and more animated conversation with the world."

This art practise is situated between art and AI, with an emphasis on artistic production, but with a strong interest in the rigorous application of an artistic idea in a computational system. The very notion of a "Dreaming Machine" is meant to highlight the weakness, rather than the strength, of AI. Self-reflection and criticism are essentially embedded in art practise, which will be reflected in this research. The dichotomy between "doing" and "just talking" discussed by Agre resonates strongly with a conception of "doing" vs "representing" which references a disconnect between the artwork and the concept. This practise aims to create software systems that are the concept, not simply manifestations of it. This concept/software unity is considered an artifact that integrates rich philosophical, cultural and technical spaces of enquiry.

[59] S. Ahson and A. Buller. *Human-Computer Systems Interaction: Backgrounds and Applications*, chapter Toward Daydreaming Machines. Springer Verlag, 2009.

The authors propose an integration of current AI methods with conceptions of mind associated with the work of early psychologists William James and Sigmund Freud. In particular, a fusion of Naur's Synaptic State Theory and Machine Psychodynamics is discussed. The purpose of this integration is to broaden the current rationally bounded conceptions of human intelligence. The "flag" computational model of cognition, ACT-R, is still bounded to rational behaviours resulting from an emotional context. There is no doubt that humans engage in "rich episodes of waking fantasy", which are not simply side effects and should be studied.

The Synaptic State Theory (SST) is a model well suited to computational implementation and considers the nervous system as composed of synapses, neurons and nodes. Synapses and neurons allow the conduction of signals while nodes sum and distribute those signals. The SST model is composed of five interacting layers: (1) The "item layer" contains nodes that are tuned to particular high level objects and concepts, which can change over time. (2) The "attention layer" allows the focus on particular senses while inhibiting others. (3) The "specious present layer" has a similar structure as the attention layer, but keeps a number of item nodes activated such that 5-10 of them are active at any one time. (4) The "sense layer" are nodes that relay sense signals from the environment. (5) The "motor layer" ties nodes to muscle control. (6) A special connection between motor and sense areas allows proprioception, which is considered the root of all emotional states. Many features of SST follow directly from William James's psychology, but provides no explanation of behaviours such as ambivalence or adventurousness.

Machine Psychodynamics (MΨD) is an approach to robotics highly marginalized by the mainstream community. "Psychodynamic" robots attempt to maximize their pleasure, which is defined as the rapid decrease of tension. Tension is the degree to which the current state of the robot diverges from its resting state. This conception of tension is not limited to tension in the body, but also includes tension resulting from divergence of internal states. As pleasure can only result from a decrease of tension, the model allows the robots to "intentionally" increase their tension through risky behaviours, such as traversing a dangerous environment without reason to do so. A third key concept in MΨD is ambivalence, where a robot can, and is expected to, hesitate in the face of conflicting options.

The fusion of these two models allows human-like behaviours. For example, if such a system had internal representations of the world, including itself, then one possible way of releasing tension could be the modification of memories that cause the tension. In Freudian terms memories could be suppressed into the subconscious, or "defence mechanisms" could be used to resolve the internal tension. Moral dilemmas could be manifest in tensions between the robot's desired reality and the reality as it stands. Just as the robot may explore the physical space in order to maximize its pleasure, it may also explore the space of its own memories, resulting in a "day-dream" where "…the robot may embellish facts, design new adventures, or even imagine completely fantastic worlds."

The question of machine consciousness can be divided into two major groups: The "hard requirement" and "modest requirement" camps. In the modest camp there are a number of criteria to determine if a machine is "conscious". The Braitenberg criterion states that a system that is not simply causal (stimulus-response) may be conscious. The Brooks criterion states that a system may be conscious if it contains no specific (hardware) location for goals, and yet can achieve them. MΨD systems meet both these requirements, and add an additional one: "the ability to achieve a state defined as pleasurable by deliberately plunging oneself into a state defined as unpleasant." The Koch criterion states that "one of the signs that a creature may be endowed with consciousness is a behaviour revealing hesitation about what to do…", which corresponds to Psychodynamic ambivalence.

Both of the models discussed in the paper are highly relevant to, and philosophically compatible with, this research. They escape from the reason and problem-solving centred aspects of AI, which are not appropriate models to implement a "machine that dreams". The SST model includes a framework in which embodiment and conceptual development can be considered. It also appears reasonably consistent with neurology. The model also includes a concise conception of emotion that is grounded in embodiment: the abstraction of proprioception. The pleasure seeking (tension releasing) behaviour of psychodynamic robots appears similar to cognitive dissonance, and provides a generator for behaviour based on homoeostasis. "Dreaming Machine" could be implemented using SST as informed by the literature in conceptual development. The notion of cognitive homoeostasis is useful as it is a

generator for behaviour (intentionality) that is not based on predetermined criteria, although does require a definition of an ideal resting state.

[60] N. Almassy, G. M. Edelman, and O. Sporns. Behavioral constraints in the development of neuronal properties: a cortical model embedded in a real-world device. *Cerebral Cortex*, 8(4):346, 1998.

"Darwin V" is the 5th project, in this series, exploring the embodiment of a neural simulation. It is meant to overcome the difficulty in determining the relation between environmental stimulus, and corresponding neural structures. "Darwin V" consists of a restricted laboratory environment that contains objects, and an autonomous robot that is able to move around the environment. The environment is painted all black and contains objects with black and white patterns. Objects with white blobs are conductive while striped objects are non-conductive. The project is composed of four major components, a visual system, a taste system, motor neurons and an ascending value system. The visual system is fed by a video camera and is composed of two subsystems: (1) A topographical layer which corresponds to the primary visual cortex and responds to low level features. (2) A higher level map that is meant to be sensitive to visual structures in a scale, orientation, and translation independent manner. The taste system is attached to a 'snout' on the robot. The snout has probes to measure the conductance of objects. Objects are 'gripped' by the snout through the use of an electromagnet. Motor neurons simply trigger pre-programed behaviours. The value system allows the high level visual structures in the visual system to trigger motor responses.

During the experiment the robot executes its default behaviour, to explore the environment, and indiscriminately grips objects. For each object gripped the conductance is registered. The conductance of the object causes the taste system to trigger either an aversive (dropping the object and turning away) or appetitive (prolonged gripping followed by moving object to the side and continuing) behaviour, depending on the conductance. Through sensor experience in the environment, the visual system becomes able to discriminate between the patterns on top of objects. The value system links the motor responses triggered by the taste to the visual recognition. Eventually the visual stimulus is associated with the taste of the object, and motor behaviours can be triggered directly by the visual system without requiring the robot to 'taste' the objects. This procedure allows the examination of both neural and behavioural variables simultaneously, which is extremely difficult with living subjects.

In one experiment the development of the visual system of a normal Darwin V is compared to the development of a Darwin V whose movement is disabled. In the latter case the visual system was fed discontinuous sensor data corresponding to the patterns on the objects, at various locations in the visual field. The authors found that the immobile Darwin V's ability to recognize objects was disrupted when compared to the mobile case. The authors conclude that the continuous movement of the robot is required for the proper development of the visual system.

Unfortunately the authors conflate the motion of the robot with continuous sensory data. They did not compare the development of the visual system of a normal Darwin V with an immobile Darwin V fed with the sensor data collected by the mobile robot. This would disambiguate the effect of continuous motion and self-generated motion. This work is significant because it suggests that an immobile sensory platform (for example an art installation) may be problematic from a development perspective.

[61] R. A. Brooks. Intelligence without representation. *Foundations of Artificial Intelligence*, 47:139–159, 1992.

Subsumption architecture, an approach to robotics control that does not depend on global representations, is discussed in this paper. This approach is developed in response to the failures of AI at the time, in particular its use of "toy worlds" and the over decomposition of functionality. Global representations "simply get in the way" when building complete systems.

For Brooks, humanity is one example of "intelligent entities" and that animals could also provide clues to intelligence. He believes that the "harder" aspect of intelligence is the embodied and adaptive relationship between an animal and its environment. The traditional focus of

AI, high level reasoning, depends on competencies that allow an animal to survive in the environment. It took much longer for evolutionary development of dynamic mobility, sensation and the "maintenance of life and reproduction" than the comparatively short development of human intelligence.

The focus on reasoning is based on an over use of abstraction in AI methodology. AI has developed such small sub-problems that they do very little on their own, and are rarely connected into complete systems. The focus on representation means that a task or world can be reduced to the point that it has little resemblance to an actual competency in an real-world environment. Abstractions are provided by the researchers in order to match our perception with that of the machine. This may not be appropriate as the sensory organs in machines bare only slight resemblance to our own. There is no guarantee that the abstractions we impose on machines resemble the internal abstractions used by us.

A core aspect of the subsumption methodology is the notion of "incremental intelligence". The aim is to build "...completely autonomous mobile agents that co-exist in the world with humans, and are seen by those humans as intelligent beings in their own right." This intelligence does not arise in a single step. The capacity of human infants does not match the capacity of adults. These creatures should have the following abilities: They must cope appropriately, and in a timely fashion, to changes in the environment. The creature must be robust in that minor changes in the world do not lead to catastrophic failure. A creature should be able to maintain multiple goals simultaneously and change their priority when opportunities arise. Finally, a creature should "do something in the world; it should have some purpose in being."

In subsumption, decomposition is based on activity, rather than function. The architecture is composed of a set of layers, running in parallel, that directly connect senses to motor actions using a network of finite state machines. Each layer enables a particular activity, and new ones are added only once the existing repertoire has been thoroughly tested in an actual environment. Layers communicate by suppressing or inhibiting the outputs of other layers. As each layer is somewhat independent, there is no central representation that models the world—the world acts as a model in itself. The lack of explicit representation allows these creatures to meet the requirements discussed above. The success of these creatures indicates that intelligence is possible without centralized explicit representations and that perhaps human intelligence could also be enabled without such representations.

The arguments that inspired subsumption architecture, the artificial and extreme decomposition of intelligence into abstract units, and the dependence on central and explicit representations, are shared with this research project. The system's conceptual development process aims to create rich linked representations that can be used to initiate dream processes, and not to enable operations in symbolic logic. The system is not intended to be a mobile robot doing practical tasks in the world, but a situated entity that collects sensory information from the world and attempt to make sense of it.

[62] R. A. Brooks, C. Breazeal, M. Marjanović, B. Scassellati, and M. M. Williamson. The Cog project: Building a humanoid robot. In *Computation for metaphors, analogy, and agents*, pages 52–87. Springer-Verlag, 1999.

Brooks discusses the development of "Cog", a humanoid robot torso meant to research aspects of embodied cognition. The development of Cog is inspired by human development, cognitive science and neuroscience. The methodology used to develop Cog rejects three central aspects of traditional artificial intelligence: (1) Monolithic models, (2) Monolithic control systems and (3) the universal applicability (general purposefulness) of systems. The emphasis is on four features of intelligence ignored by the traditional AI of the time: (1) Development, the development of skills and abilities over time as a human child might. (2) Social interaction, the ability to get assistance from, and collaborate with, other agents. (3) Physical interaction, in place of a monolithic model, the world is considered a tool for the manipulation of knowledge. (4) The integration of sensorimotor signals with cognitive processes.

At the time of writing Cog had developed a number of abilities. At the core of future development is Cog's ability to track the gaze of a care-giver in order to form joint-attention. This is ability is believed to be the basis of early infant-caregiver communication in learning. Cog exhibits human-like arm movements based on an integrated coupling between intentional movement and the physical characteristics of the arms (for example capitalizing on pendulum motion, etc.). Abilities in development include various visual motor routines: the ability to saccade to objects, track moving objects using smooth pursuit, match of head and neck movements to keep objects in view, point to visual targets, recognize joint attention through face and eye finding, imitate head nods, and the regulation of actions through facial expression.

Cog is meant to explore embodied intelligence informed by knowledge of human development. The computational implementation of cognitive and developmental theories provides a valuable starting point in embodied cognition. Brooks rejection of monolithic models may contradict the importance of explicit representation at the core of human development. Additionally, Cog's planned ability to learn is dependent on a infant-caregiver relationship, which may not be appropriate for a public media art installation.

[63] H. Q. Chong, A. H. Tan, and G. W. Ng. Integrated cognitive architectures: a survey. *Artificial Intelligence Review*, 28:103–130, 2007. 10.1007/s10462-009-9094-9.

Integrated cognitive architectures are systems that model human behaviour at the cognitive level. This survey covers six cognitive architectures: State, operator, and result (Soar), Adaptive control of thought-rational (ACT-R), ICARUS, Belief-Desire-Intention (BDI), Subsumption Architecture and Connectionist learning with adaptive rule induction on-line (CLARION). Soar is one of the oldest and most developed architectures. It is based on the "physical symbolic hypothesis" and is rooted in symbolic approaches to AI. The system has a single unified working memory which pulls from long term procedural, declarative, and episodic memory systems. Soar supports a hierarchy of goals in order to decompose a single goal into sub-goals. Goals are attained through the application of selected operations (actions), and by means-ends analysis. Soar uses reinforcement learning to reward successful applications of operations toward goals.

ACT-R is highly oriented towards biological plausibility. The mapping between system modules and brain regions allows the system to be validated from brain imaging data. The architecture is composed of modules that perform visual perception, motor control, declarative memory and goal oriented functions. Each module interfaces with the "central production system" (CPS) via independent buffers (short-term memory) to limit the flow of information from the modules' long-term memory. Like Soar, ACT-R uses symbolic representations of knowledge. Goals are explicitly handled by the goal module. Problem solving in ACT-R is enabled by a Bayesian framework. ACT-R can learn by linking existing actions into larger meta-actions that can be applied more efficiently.

The ICARUS architecture is composed of four components. The "perceptual buffer" corresponds to short-term perceptual memory. Both long-term and short-term conceptual (declarative) memory are managed by the "concept memory" module. The "skill memory" module contains a long-term store of skills, as well as short-term storage for the skill being currently implemented. Actions in the world are triggered in the motor buffer by the skill memory module. Goals in ICARUS are stored in a stack and hierarchically organized. Problem solving is accomplished by means-ends analysis.

BDI is based on Dennet's theory of intentional systems and the theory of human practical reasoning. The major components of the BDI architecture are belief, desire and intention. Beliefs are facts about the world and inference rules for the construction of new beliefs. Plans are sequences of actions used to perform intentions. They are composed of a body (actions and procedures) and an invocation condition (prerequisites for the plan to executed). Desires are the aims of the agent but unlike goals they can conflict. Intentions are the actions the agent is committed to in order to achieve its desires. BDI agents make use of means-ends analysis in problem solving. Reinforcement learning is a recent addition to the architecture.

Subsumption is a robotics oriented architecture where behaviours are built from the bottom-up without the use of declarative knowledge. The bottom-up approach involves building simple behaviours that can be inhibited by higher level layers. All layers have access to sensor input and motor action. Each layer is executed separately, so the loss of one layer/behaviour does not effect other layers/behaviours.

CLARION is rooted in connectionist AI, and makes use of both symbolic and sub-symbolic knowledge representation. The architecture can be decomposed into two levels. The Associative Memory Networks (AMN) is the bottom level and learns procedural knowledge using reinforcement learning. The General Knowledge Store (GKS) is the symbolic counterpart of the AMN. It creates declarative representations from the procedural knowledge contained in the AMN. Both levels run in parallel. When an action is to be selected, it is based on a weighed sum from the results of both sub-systems. Two additional subsystems are the motivational subsystem, which guides behaviour by forming goals; and the meta-cognitive subsystem, which manages the interaction between other subsystems in order to generate structured behaviour.

All of the cognitive architectures discussed above are oriented toward symbolic knowledge representation with the exception of CLARION which uses a hybrid of symbolic and sub-symbolic representation. Only ACT-R is neurobiologically oriented, with a specific mapping between modules in the architecture and brain regions. Cognitive architectures are high level computational models of human behaviour, making them relevant to the development of a machine that dreams. Their emphasis on reason and problem solving is problematic as the role of dreaming in these areas is not well understood. Of particular value is the method CLARION uses to extract high level symbolic representation from knowledge learned in a connectionist system. The method is highly relevant to the development of concepts (explicit representations) which arise from low-level sensory (sub-symbolic) correlations in infant development.

[64] J. Hawkins and D. George. Hierarchical temporal memory: Concepts, theory and terminology. *Whitepaper, Numenta Inc*, 2006.

Hierarchical Temporal Memory (HTM) is an approach to machine learning inspired by the human neocortex and is technically similar to a Bayesian Network. An HTM learns patterns in both space and time, in a hierarchical fashion. Inputs are fed into a large number of leaf nodes at the bottom, where each node receives a subset of the sensory space. Multiple leaf nodes feed higher level nodes, which are sensitive to structure in multiple areas corresponding to the lower level nodes that feed it. HTMs are able to "discover" "causes" that underlie low-level sensor data and create hierarchical representations of causes. They can discover causes even when fed novel sensory data, and are predictive—in theory they can "imagine" future sensor states. HTMs can also be tuned using an attention mechanism.

Each node executes a common algorithm that learns temporal and spatial patterns. Learning is the process thereby causes in the world are discovered through correlations in the state of sensors. Each node stores a fixed number of patterns in both space and time. The learning algorithm proceeds as follows: (1) A fixed number of the most common spatial patterns (quantization points) are stored in each node. A distance function is used to measure the similarity of input patterns to stored patterns, which are changed gradually to match the most commonly received patterns. (2) For each input, a probability distribution is generated that reflects the likelihood that the input pattern matches each of the stored patterns. (3) The changes of these distributions over time are tracked and a fixed number of the most common sequences are stored. Each sequence represents the nodes "belief"—the probability that the current input matches a known cause. (4) Units send beliefs to their parents, where they are treated as a new input patterns. The next level discovers causes of causes (spatial patterns of sequences of spatial patterns of sequences), and so on up the hierarchy. High level patterns are then correlations between very large numbers of input patterns.

HTMs are significant to this work because they use temporal correlation to learn associations between patterns that do not have similar features. This is similar to Mandler's conception of Image-Schema theory where the meaning of objects infants see in the world is due to their behaviour in time, and not their object-centric features. Additionally, HTMs can be used in an unsupervised fashion, and attempt to learn the structure, in both space and

time, of arbitrary sense data. Unfortunately they learn offline, are not program independent, and their structure is not self-organizing, but has to be designed for a particular application.

[65] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

Kohonen provides a survey of SOM methods and possible applications. The SOM is an unsupervised artificial neural network (ANN). The result of a SOM is a map where regions represent "a particular domain of input patterns". Although considered an ANN, the design emphasizes the technical over the biological. The SOM is expected to be similar to the human cerebral cortex where regions are tuned to varying modalities and arranged by features of stimuli.

The SOM is a development on top of classic competitive learning methods, in particular "Vector Quantization" (VQ). In VQ a number of code-books are used to represent the distribution of inputs. These code-books may initially be random, or determined from an analysis of the distribution. The code-book most similar, in terms of a euclidean distance measure, to the input is the "winner". The value of the winner is made more similar to the input, according to a learning rate that decreases monotonically over time. Maps such as these have not been found in brains and tend to show self-organization only on a local scale and are often low in dimensionality.

The SOM is an extension of VQ methods to ensure that a global self-organized structure forms during training such that it reflects the topology of the set of input patterns. The SOM concentrates on the spatial arrangement of code-books, which are tuned from input patterns. The algorithm proceeds as follows: (1) The selection of the best matching unit (BMU) is accomplished by determining which code-book is most similar to the current input—the code-book with the minimum euclidean distance. (2) Code-books are updated only in a subset of the map. This area is called a "neighbourhood", is centred around the BMU, and allows global self-organization of the map. The size of the neighbourhood is large at the start of training and monotonically decreases over time. The initial large neighbourhood serves to rough out the topological arrangement of code-books, which are increasingly refined by a smaller and smaller neighbourhood. The influence of learning within the neighbourhood can be determined using a hard limit, a linear ramp or a Gaussian function. Units within the neighbourhood learn using the same function as in VQ.

For a SOM to converge at a map that properly reflects input data, the number of iterations is crucial. A rule of thumb is to train with 500 times the number of units in the map. The neighbourhood and learning rates should not change for the first 1000 and final 1000 iterations. The size of the initial neighbourhood should cover the whole network for proper global organization. Input values fed to the SOM should be high level features, such as histograms and not low-level features, such as pixels.

The SOM has been applied to the hierarchical clustering, classification (although "Learning Vector Quantizers" (LVQs)—SOM variants—are required), speech recognition, and semantic mapping. In semantic maps the main issue is the distance measure, which is not clear for symbolic values. The authors consider the context around a symbol in this case. Two query symbols surrounded by the same symbols, the same context, are considered similar. The number of symbols shared between contexts shows increasing similarity. Using this method a SOM was able to differentiate between nouns, verbs and adjectives from simple sentences. Other possible applications include robot control, synthesis, compression, optimization, etc..

The SOM was the central mechanism to organize memory in previous work on "Memory Association Machine" and early "Dreaming Machines". The method fit the requirements of the project by being unsupervised and not depending on many assumptions regarding input data. This method is not considered for current research because it does not train well in a continuous fashion, and therefore requires all input patterns to be stored. Growing and hierarchical SOMs are possible, but these also require all patterns to be stored and repeated training. Another self-organizing method is required, or a reliable extension that allows continuous training, without storing all patterns.

[66] S. J. Russell and P. Norvig. *Artificial intelligence: a modern approach*, chapter 13 Quantifying Uncertainty. Prentice hall, 2009.

This chapter provides the mathematical foundations of probability theory in the service of designing artificial agents meant to operate in uncertain environments. The uncertainty of an environment can be due the agent's ability to observe only part of the environment at a time, non-determinism or complexity. Non-probabilistic logic is inappropriate in uncertain situations because of three reasons: (1) An uncertain situation could include an infinite number of exceptions to any rule. (2) There may be no complete theory to cover all cases. (3) It is impossible to test every single case, even if there is a complete theory. Probabilities are the likelihood that a particular world in the sample space fits the current situation. A joint probability distribution is a table of probabilities for each variable, in terms of related variables. A Full Joint Probability Distribution (FJPD) contains the probabilities for every possible world for every variable. Probabilistic Inference makes use of the FJPD in order to determine what the probability of X is, given Y. Bayes' rule can be used to calculate the unknown probability of a variable from the known probabilities of related variables. The FJPD does not scale linearly, conditional independence can be used to break the FJPD into multiple smaller distributions that are independent.

The system produced during this research program is expected to inhabit a complex and real world environment, which certainly qualifies as uncertain. The long term that the artwork is expected to be installed makes statistical learning methods highly appropriate, both in terms of the amount of statistical data that can be collected, and in terms of the compactness of the stored probability distributions.

[67] S. J. Russell and P. Norvig. *Artificial intelligence: a modern approach*, chapter 14 Probabilistic Reasoning. Prentice hall, 2009.

This chapter discusses mathematical methods for representing uncertain knowledge and doing inference in an uncertain domain. The Bayesian Network (BN) is a directed acyclic graph that represents a Full Joint Probability Distribution (FJPD) for a certain domain. Each node in the BN is a random variable, and each is associated with a Conditional Probability Table (CPT) given each nodes parents. The same kind of inference that can be done with a FJPD can be done with a BN, but the BN represents the distribution in a more compact fashion. The Chain Rule is useful in determining if a BN is a good representation of a domain by insuring that each node is conditionally independent of other predecessors in the node ordering, given its parents. This method constructs BNs that contain no redundancy, unlike FJPDs, and cannot break the axioms of probability. BNs can contain Boolean or continuous variables, or both in the case of hybrid BNs. As continuous variables may have an infinite number of possible values, the exact distribution cannot be determined and canonical distributions are often used.

Exact inference in BNs is possible, but in the general case it is intractable. Methods for exact inference include: (1) Inference By Enumeration, which inefficiently sums of the products of all the conditional probabilities in the network. (2) The Variable Elimination Algorithm, which reduces redundancy by caching segments of the calculation. In a singly connected network (polytree) the time and space complexity grow linearly with the number of nodes. In a multiply connected network the complexity increases exponentially and are #P-hard. A multiply connected network can be converted to a polytree using a clustering method to group multiple nodes into single nodes with larger CPTs. Special inference algorithms are required for this case.

Due to the intractability of exact inference in BNs, a number of methods for approximate inference have been developed: (1) Direct sampling methods set variables in the network and record the results. The results can be used to approximate the exact probabilities. Two such methods include rejection sampling, where samples are rejected if they don't match evidence, and likelihood weighting, where events are weighted based on their likelihood. (2) Markov Chain methods use random values to explore the network's state space. The number of iterations spent in each state is proportional to the posterior probability. Methods in this

family include Markov Chain Monte Carlo and Gibbs Sampling. In these methods, as the number of samples increase the approximate solution approaches the exact solution.

BNs are powerful representations but in some cases they can contain redundant information. First order probability models are an effort to combine the expressiveness of first order logic and the efficiency of probability models. Alternatives to BNs include: Rule-based systems (logic) for uncertain reasoning, Dempster-Shafer Theory, and fuzzy methods.

The statistical methods discussed in this chapter may be appropriate to this research program because the long term aspect of the artwork requires some method of learning from patterns in the world without those patterns being stored. A continuously updating distribution of occurrences in place of actual patterns is very attractive. The relation between the learning of a distribution with a label and concept generation is unclear at this stage.

[68] B. Scassellati. Theory of mind for a humanoid robot. *Autonomous Robots*, 12(1):13–24, 2002.

"Theory of mind" (ToM) is the concept of the mind of another. It occurs when an agent is able to differentiate between their beliefs, desires and goals (BDG) and those of others. It would be required in order for a humanoid robot to act socially in a "natural" manner. A ToM is composed of the abilities that allow an agent to make this distinction, and is only valid in a intentional or goal-directed framework. The differentiation of BDG is instrumental in the development of self-recognition, and perhaps in grounding meaning in language acquisition, imagination and creative play. This paper describes two leading theories of mind: Leslie (1984) and Baron-Cohen (1995), and discusses implementations of related abilities in "Cog" (a humanoid robot).

Leslie's model of theory of mind is based on the representation of causal events, which is also used to develop a theory of object mechanics. The model is comprised of three classes of events, based on causal structure, that are managed by three modules. The "Mechanical Agency" class corresponds to the "Theory of Body" module, and is the first to develop. It describes how infants understand the physical movement of inanimate objects. There are two types of visual input, three dimensional object centric processing, and object-independent motion based processing. Leslie assumes that the latter is innate, but it has been argued that it may arise very early in development. "Actional Agency" events correspond to the "Theory of Mind Module 1", and develops second. This module represents the infant's understanding of intentions and goals of other agents, example representations include: approach, avoid, and escape. Infants begin to focus on the movement of eyes, rather than whole objects. "Attitudinal Agency" corresponds to "Theory of Mind Module 2" and is the last to develop. This module deals with the representation of attitudes and beliefs of other agents. It makes use of meta-representation and "...allows truth properties of a statement to be based on mental states rather than observable stimuli." This level of development is required for the infant to conceptualize beliefs that contradict ones own, and may be the basis of pretend play.

Baron-Cohen's model of theory of mind depends on two forms of perceptual input: the "Intentionality Detector" (ID) interprets self-propelled motion in terms of goals and desires, while the "Eye Direction Detector" (EDD) is tuned to all eye-like shapes and determines the direction of gaze of other agents. The EDD is composed of three sub-modules: one detects eye-like shapes in the visual field, another determines if the gaze is directed towards the infant, and the third interprets gaze direction as a "perceptual state". The "Shared Attention Mechanism" allows infants to glean the goals of other agents through their eye direction. The combination of EDD and ID representations allows the infant to conceptualize statements such as: John sees that I want the toy. The "Theory of Mind Mechanism" represents the mental states of other agents and ties them into a coherent whole. In this module statements such as "John believes it is raining" are possible to cognise, even if the agent knows its not raining. This model is supported by developmental evidence, and applied to the study of autism spectrum disorders and non-humans.

A ToM for a robot would allow human-robot interaction to operate similarly to human-human interaction. Robots could learn like children, which would allow them to acquire skills from care-givers who do not possess specialized robotics knowledge. The interpretation of the emotional state of humans would allow a robot to act appropriately in more situations. An implementation of a synthetic ToM could validate proposed models.

The paper describes a computational system that combines elements of both Baron-Cohen and Leslie models. These mechanisms are implemented and evaluated in the Cog robot. Cog has a number of pre-attentive visual routines providing colour saliency, motion and skin colour detection. The attentive visual system allows the robot to saccade to areas containing colour contrast, faces, or motion. Cog is able to find faces and eyes in the visual field, and saccade to them. Cog is able to discriminate between animate and inanimate objects, which is determined through an analysis of motion vectors. Future work on Cog includes gaze following, extension of gaze following using deictic gestures, ie. pointing, and further refinements to animate-inanimate distinction.

The gap between theory and implementation is clearly large. On one side is the fairly complete description of theory of mind models, while the other side is a description of low-level visual computing mechanisms that bare only some resemblance to ToM. Leslie's model is consistent with current evidence supporting object and location/movement based visual processing. It is also consistent with Mandler's image-schema theory of conceptual development. Baron-Cohen's model appears more computationally feasible and is used more in Cog's development. Theory of mind is an example of a very high level concept that is formed through development. How it arises, and its possible relation to imagination and play, is relevant to this research. The computational techniques used to implement the low-level perceptual precursors to ToM could be used in this project.

[69] R. Sun. CONSYDERR: a two-level hybrid architecture for structuring knowledge for commonsense reasoning. In *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*, volume 3, pages 1475–1480, 1994.

CONSYDERR (Connectionist System with Dual-representation for Evidential Robust Reasoning) is a highly parallel system that includes explicit levels representing concepts and the features of those concepts. It aims to reflect the general applicability and robustness to uncertainty of human knowledge application. Knowledge is encoded as simple and compound "Horn statements" that represent propositions and rules. The architecture's two levels are each composed of a set of nodes that are linked. At the "concept level", propositions are nodes and rules link nodes representing antecedents and consequents. At the "micro-feature level", the features of concepts are represented by sets of nodes, linked to the concept level. The activation of concepts results in the activation of relevant micro-features, and vice versa. The rule links between concepts are "diffusely" replicated in links between their sets of respective micro-features. The system explicitly deals with two types of vagueness: (1) Each rule is applied to a degree, due to the weighted sum of antecedents. (2) Two concepts can be considered similar, due to the degree to which their micro-features overlap. This is handled by a three phase procedure: (1) In the top-down phase, the activation of micro-features results from the maximum weighted activation from linked concepts. (2) In the settling phase, for both levels, the activation of nodes is a consequence of the weighted sum of antecedent activations in the same level. (3) In the bottom-up phase, the activation of a concept node results from the weighted sum of its associated micro-feature nodes, but only if that activation is greater than its initial activation. The system is able to exhibit three types of commonsense reasoning: (1) Similarity allows the application of rules that don't exactly match the current case. (2) Inference results from the application of rules. (3) Inheritance is possible through inference based on superclass/subclass relations by an analysis of the concepts' micro-features.

This proposal appears to be a system of relating two connectionist systems, one corresponding to concepts, which may be hierarchical, and the other corresponding to the features of those concepts. If we consider these features as stimulus, then concepts can result from perception, although it appears in this case that features are derived from concepts. Since concepts can lead to features, and features can lead to concepts this is a natural consideration of concept generation, and mental imagery (generation of micro-features based on the activation of a concept). Unfortunately the conceptual system is not grown from stimulus

but is imputed through knowledge statements. It is also unclear what the atoms of perception are, and how they would be generated by concepts and still form readable images. The development literature, Mandler in particular, could provide methods to grow concepts from micro-features. Additionally the concept/feature split resembles Zhang's interpretation of "symbol" and "representation" brain hemispheres.

[70] A. Turing. *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, Plus the Secrets of Enigma*, chapter Computing Machinery and Intelligence (1950). Oxford University Press, USA, 2004.

This seminal paper describes Alan Turing's position on the possibility of machine "thinking". Rather than dealing with the poorly defined question "Can machines think?", Turing proposes a test. The "imitation game" involves three people, a man, a woman, and an interrogator. The interrogator's role is to determine which is a man and which is a woman, without any physical cues, and through a teletype conversation where he or she is free to ask any question. The man's role is to trick the interrogator by presenting himself as a woman, while the woman presents herself truthfully. Turing proposes that the man is replaced by a machine. The interrogator then must determine which is a machine and which is not. Turing considers this test more useful than the original question, and discusses nine arguments against it.

(1) The theological argument states that only people with souls can think, and machines don't have souls. If people are God's instruments, and if we make a machine that passes the test, then God has given that machine a soul. (2) The idea of a thinking machine is troublesome to some, and therefore we should not ask the question. (3) Machines are limited, and as such cannot approach human behaviour. A machine cannot have the "correct" answer to all possible questions, but neither would a person. (4) A thinking machine should be able to create, have emotion, and be self-aware. This argument falls into solipsism, as the only way to for the machine to pass the test, would be for it to be aware of itself doing so. (5) There will always be tasks that one could point to that a machine could not accomplish. These arguments are often veiled references to argument #4. (6) Lovelace states that machines can't originate, but can only do what we know how to direct them to do. Turing breaks down the idea of "originality" where it could result from the growth of a planted seed, or the effect of following known principals. (7) The brain is not a discrete state machine and therefore a machine can't emulate it. (8) There is no rulebook that tells a man what to do in any situation, so a machine that is bound by laws is not like a man. (9) Turing states that evidence for telepathy is 'overwhelming', and perhaps would easily allow the interrogator to pick out the machine. The only method of avoiding this would be the use of a "telepathy proof room" in the test.

Turing proposes the idea of a learning machine, as the core problem in making a machine to pass the test is more a programming challenge than an engineering one. What if a machine was like a child and could learn, rather than be programmed? Turing proposes learning approaches similar to evolutionary algorithms, reinforcement learning, and logical inference systems. "Intelligent behaviour presumably consists in a departure from the completely disciplined behaviour involved in computation, but a rather slight one, which does not give rise to random behaviour, or to pointless repetitive loops." This quote sets up the challenge nicely, intelligence is not simply reasoning, not random, nor is composed of useless repetitive loops.

The question "can machines think" is analogous to the question asked in this research project, "can a machine dream?" Turing re-frames the question with a test, that shows that a machine can appear to be a thinking person. The validity of this research project is not in the similarity between the machine's and a human's 'dreams', but in the application of current knowledge of cognitive processes. The arguments and responses Turing discusses against a thinking machine equally apply to a "dreaming machine".

7 Related Work

[71] B. D. R. Bogart. Memory association machine: An account of the realization and interpretation of an autonomous responsive site-specific artwork. Master's thesis, Simon Fraser University, 2008.

This thesis is an account of the realization and interpretation of the autonomous responsive electronic media artwork "Memory Association Machine" (MAM). Realization and interpretation are components of the creative process that braids conceptual, site-specific, electronic media art and artificial intelligence practises. The meaning of MAM is dependent on its unique location in space and time. MAM relates itself to its context using three primary processes: perception, the integration of sense data into a field of experience, and the free-association through that field. MAM perceives through a video camera, integrates using a Kohonen Self-Organizing Map, and free-associates through an implementation of Liane M. Gabora's model of memory and creativity. These processes are as important as MAM's physical appearance, are composed of computational elements, and allow the system to respond to context autonomously.

This research follow from my masters work in that the "dreaming machine" is expected to be a site-specific artwork in the same sense as MAM: It is meant to automatically "relate to its context" and its generative output is meant to be surprising. Both are art-as-research projects.

[72] Z. Byers, M. Dixon, W. D. Smart, and C. M. Grimm. Say cheese! Experiences with a robot photographer. *AI magazine*, 25(3):37, 2004.

The authors describe a robotic photographer that is meant to capture documentation images of people at public events, such as conferences and weddings. The system uses computer vision to frame and compose each photo. The robot is human sized and intended for unaltered interior environments. Its vision system is made up of a still and video camera, which are mounted on a pan/tilt unit. Navigation is accomplished using a laser range-finder. The robot randomly explores a space looking for a photographic opportunity. An opportunity arises when the video camera detects a face, the positions of which are used in framing and composition.

The robot was constructed to explore issues in long-term autonomy, autonomous navigation and robot-human interaction. Photography was chosen so that the robot could interact with people without specialized knowledge. Face detection is based on skin-tone analysis in the video image. The location of faces in the video image is combined with range-finder data to move the robot into position. The quality of the photographic opportunity is determined by the distance between camera and subject, the occlusion of faces, the symmetrical arrangement of faces, movement within frame and the accessibility of the potential subject. Once an opportunity has been found, the robot moves toward the goal, while simultaneously avoiding obstacles and recalculating the potential of the opportunity. If navigation to the target is too difficult then the opportunity is abandoned. There is also a random mode where the robot moves randomly and constantly looks for an opportunity. This mode is most useful in crowds where accurate navigation is difficult. The framing of the image is determined based on aesthetic criteria: (1) According to the rule of thirds, faces should be 1/3 or 2/3 from frame edges. (2) The area of the face should be 1/3 of the width or height of the frame. (3) Faces should not appear in the middle of the frame, and (4) Faces should not be bisected by the frame. Images captured by the system are displayed at a "viewing station" where they can be deleted and rated. Image ratings could be used to help the photographer determine what is a good composition.

There are few creatively oriented projects where robots capture images of their environment. Although the project proposed in this research is not mobile, it does capture images of its environment in an autonomous fashion. Research in infant development emphasize the importance of movement over object features. As such a video camera will likely be used in this project, over a computer controlled still camera. In order to keep the perceptual system as simple as possible standard computer vision methods will be used. A static camera may be required in order to use background subtraction for segmentation.

[73] H. Cohen. What is an image? In *Proceedings of IJCAI*, 1979.

AARON is an expert system that paints in a distinctive human-like style similar to Cohen's. AARON is implemented in Lisp and is a large rule-based system in which weighted and

constrained randomness provides variation. For Cohen, the process of image-making is a cognitive and inferential process. AARON paints from internal models of the world, which are explicitly defined by Cohen. The purpose of the project is not to make a painting machine, but "... to understand more about the nature of art-making processes than the making of art itself allows... ".

AARON is generative; No interactive input drives its internal process, only the rules specified by Cohen. AARON does not reflect on its results, and has no sensory system to "see" that which it paints. AARON's style is a manifestation of Cohen's interest in exploring a particular property of free-hard drawing: what is the minimum number of markings needed to make a representational image? According to Cohen, AARON is not an artist but is intended to be an image-maker operating in an art context. The resulting drawings and paintings are meant to be seen as original and of sufficient quality to appear as artistic works to a sophisticated audience, whose feedback is a component of the development methodology.

AARON shifts between multiple layers of attention, moving between high level concepts and low-level formal details of the image. These levels are organized in a hierarchy that reflects AARON's architecture. The top most structure is the "Artwork" module, which manages the overall structure of the image, and defines the start and end of the drawing process. The "Mapping and Planning" module places drawing components (figures) in available space on the canvas, and defines the drawing process for individual figures. Figures are composed of multiple "developments". In each development the third layer, "Lines and Sectors", returns to planning for further instruction. Developments may consist of a number of lines, which are the result of random values constrained by rules provided by higher levels. Lines are used to compose curves that represent pen movement. Execution oscillates between these layers as AARON draws. The next two levels depend on the selected output of the system: AARON creates images both on screen, and also using a robot pen to draw on paper and canvas. The robot version of AARON has a "Movement Control" module that controls the movement of a pen. The curves created by the "Lines and Sectors" module are directly mapped to pen parameters. An additional module "Feedback" reports the pen's relative position back to the program.

For Cohen, image making and image reading are both enabled by cognitive processes. These cognitive processes are shared between maker and reader and bind the two in a "transaction". The intended meaning of an artwork plays a subordinate role to the meaning within the structure of the image. Meaning is in the form over the content. AARON is possible because the minimum requirements for constructing meaning does not include the intention to communicate, but a set of cognitive processes. Cohen identifies three cognitive processes that serve as a starting points in image-making: the ability to differentiate between figure and ground, open and closed forms and inside and outside. Art, as process rather than artifact, is universal because it is composed of cognitive processes that we all share.

AARON is one of the first and most successful examples of a creative machine. Two features are most relevant to this research project. (1) The central driving force behind AARON is the desire to understand the cognitive processes that enable image-making. These processes are not explored through a scientific methodology, but an artistic process of study and exhibition. (2) The hierarchical structure of AARON reflects a distinction between conceptual levels and the underlying sensorimotor mechanisms, which is relevant to the importance of concept generation in this research. Cohen's conception of art and art practise are strongly aligned with this work.

[74] H. Cohen. The further exploits of aaron, painter. *Stanford Humanities Review*, 4:141–158, 1995.

This paper describes an early screen-based version of AARON, a creative machine that draws and paints in a human-like style. AARON paints from internal models of figures that are hard-coded as a set of constraints. AARON has few of these models. Adding more was considered but Cohen wanted to see what AARON could do with a small vocabulary of objects. AARON's models are simplistic, three dimensional and derived from scanning points in medical illustrations of the human skeleton. The simplicity of models rejects a photo-realistic

aesthetic. Although models are three dimensional, they are stacked on a two dimensional image-plane without perspective. The version of AARON discussed in this paper is intended to produce figurative portraits in colour. Previous versions of AARON would draw in black and Cohen would paint them in colour. AARON's colouring system is an explicit rejection of the computer graphics of the day. AARON does not use a flood-fill algorithm to colour areas of an image, but uses a more painterly system where non-uniform colour blobs inaccurately fill drawn regions. In this paper Cohen explains why AARON is not an artist, but only an image-maker: the construction of art requires self-awareness, and since computer programs cannot be self-aware they cannot be artists.

This paper provides detail of the development of AARON through its earlier revisions. These versions of AARON use a robot turtle holding a pen that draws on large pieces of white paper. The cognitive orientation of AARON is relevant to this project, which aims to make use of explicit cognitive models devised by Cohen. Cohen is a significant precedent for the development of computational and cognitively oriented creative machines to explore an artistic enquiry.

[75] A. Franco. Controlled dream machine. Master's thesis, University of Plymouth, 2007.

Anaisa Franco produced "Controlled Dream Machine" during her Master of Arts studies at the University of Plymouth. The artwork consists of a pair of robotic legs, outfitted with LEDs—which correspond to a fictive character's body—and a projection of dreams—which result from the character's "subconscious". These two components pass information (behaviours in her language) back and forth, "... between consciousness (symptoms of the body) and unconscious (mind and dreams)". The project is situated in theory of art, science, philosophy and psychology.

The artwork is meant to be a whole self, including consciousness, body, self and unconsciousness. The conception of dreams used in the work is Freudian, where dreams result from the unconscious collection of repressed ideas, wishes and desires. Dreams are "fragmented narrative episodes" manifest as animations and videos sequences, framed as spaces, which are produced by the artist. These dreams are organized into five sections, where each corresponds with behaviours manifested by the legs: (1) All dreams start in "sensations", where the robot legs walk and suddenly stop as the dream begins. From this state any of the other dream categories can be initiated, and can lead from one to another. (2) In "traumas" the legs shake and LEDs blink. (3) Traumas can often lead to "nightmares" where legs shake and LEDs blink continuously. (4) In "Memories" the legs walk and stop as they remember and LEDs change to white. (5) In "collective unconsciousness" the legs walk, and then stop and shake suddenly. The selection of dream spaces is triggered by a recording of neural activity in embryonic rat neurons, cultured in an electrode array. As these cells grow and interconnect, their activity drives the artwork's dream process.

"Controlled Dream Machine" is one of the few electronic media art projects that aims to create a machine that has dreams. Despite the use of "wetware", the conception of dreaming is not situated in cognitive, nor neurological theory. Dreams are time-based artworks, produced by the artist, and represent aspects of Freud's conception of the unconscious. The communication between the neuron cells and the artwork is one way, and apparently not real-time, as they are not effected by the dreams. The complexity of the data produced by the neurons is filtered out in their triggering of 5 possible animation sequences, and the mapping is not discussed. The emphasis on the work, and artistic contextualization, appears to emphasize the robotic and physical aspect. Franco's work contrasts highly with this research, which emphasizes a rigorous connection between technical practise (AI and computation), scientific knowledge (cognitive science and neurology) and artistic concept (site-specific, conceptual and generative).

[76] N. Kamal, L. Tsou, A. Al Hajri, and S. Fels. DreamThrower: Creating, Throwing and Catching Dreams for Collaborative Dream Sharing. In Hyun Yang, Rainer Malaka, Junichi Hoshino, and Jung Han, editors, *Entertainment Computing - ICEC 2010*, volume 6243 of *Lecture Notes in Computer Science*, pages 20–31. Springer Berlin / Heidelberg, 2010.

"DreamThrower" is a system that is meant to allow people to influence and share dream experiences, and follows the "Dream Catcher" metaphor. The concept is that stimulus provided during sleep effects the dream in a particular manner such that the same stimulus, given to another person, results in a similar dream.

The system consists of a "eye mask" that detects REM sleep, provides the stimulus to alter the dream, and a final cue to alert the user when to wake up in order to report on the dream. The second component is a social network where dream reports, and stimuli, are recorded and shared. The act of throwing a dream is sending a stimulus to another user in order to initiate a similar dream. There are a number of other projects that use technological devices to alter dreams or alert the dreamer of REM sleep. The authors' pilot study concluded that while the REM sleep detection worked, the effect of stimulus on dreaming was inconclusive. Future work emphasizes the social network component of the project, not tested in the pilot study, to see if the act of socializing about dreams has a greater effect on dream content.

DreamThrower is among a number of technological devices meant to alter the dream state of a person. An overlap between the DreamThrower and this research is the question as to what the effect of external waking stimulus is on dream content. If dreams are randomly initiated then there should be little effect on waking stimulus on dream content. If we consider dreams as generated from concepts, then any concept could be triggered, allowing a large variance of dream content. As dream content does appear to correlate with features of waking experience, it seems unlikely that they would be random, and perception oriented. This may point to dreams as initiated by emergent bursts of activity in brain regions that are still slightly active from waking experience. This is highly relevant to this project as it would allow a connection between waking stimulus and dream content, and is also computationally feasible.

[77] G. Legrady and T. Honkela. Pockets Full of Memories: an interactive museum installation. *Visual Communication*, 1(2):163–169, 2002.

"Pockets Full of Memories" (PFOM) is an interactive installation conceived by George Legrady and produced in collaboration with Timo Honkela. The project is meant to explore issues of archive and memory. The installation consists of a number of scanning kiosks. Participants are invited to place objects on the scanners, and answer a questionnaire regarding their attributes. The image of the contributed object is associated with the data from the questionnaire and fed into a Self-Organizing Map (SOM). The SOM component was developed by Honkela, and organizes the meta-data associated with each contribution into a 2D grid by similarity. A large grid is projected on a wall beyond the kiosks. The SOM determines in which grid cell a particular contribution is placed.

Participants submitted a wide variety of objects, from personal electronics to scans of body parts. During the installation over 3300 objects were submitted to the system. In order to link the projected image to the most recent contributions, a subset of the entire database of objects, where items are weighted by the time of submission, is fed to the SOM. The installation also involves a web component, where all the objects could be viewed, and visitors are encouraged to submit stories linked to objects in the database.

The visual presentation of the SOM in PFOM highlights the tension between the visual appearance of the object, and the answers to the questionnaire that define the meaning of the objects. PFOM is one of the few installation projects that makes use of the SOM. This use links the project with this research that follows from SOM enabled projects such as "Memory Association Machine", "Dreaming Machine", and "Self-Organized Landscapes".

[78] S. Penny. Embodied Cultural Agents: at the intersection of robotics, cognitive science and interactive art. *Socially Intelligent Agents*, pages 103–105, 1997.

"Petit Mal" is a mobile robot artwork, created by Simon Penny, that is meant to be "truly autonomous", "nimble" and have "charm". It can sense and explore is architectural context and is pursued and related to people. The work is meant to give an impression of intelligence, but not intended to be an artificially intelligent device. Interactive art is expanded, in terms of the "bandwidth" of interaction, by this "actor in social space". The design rejects biomorphic,

anthropomorphic and zoomorphic tendencies in order to focus the viewer's attention on its behaviours, rather than on its appearance.

The robot's overall design is minimal. It consists of a two wheel drive for mobility, and two pendulums for balance and stability. The angle between the two pendulums is measured by an accelerometer that reflects the robots proprioceptive sense of balance. The physical interaction of these two pendulums, in particular the inner one, results in a not quite stable design and exhibits "unpredictability". The robot has no memory and is intentionally "anti-optimized", as optimization simply reduces the number of possible behaviours a system can exhibit. In order to maximize expressiveness and personality the robot "...teeters on the threshold between functioning and non-functioning."

Penny's conception of technical practice considers robots and computers "full of cultural meaning" which contributes to rich and varied interpretations of their behaviour. The viewer sees abilities in the robot's behaviour that it does not actually posses because the system capitalizes on cultural metaphors. A robot can only appear as an agent if it triggers the association of cultural notions in the "user". Petit Mal becomes social because the user is required to adapt to its limitations, and is therefore engaged in a social relation.

Centrally, Penny is concerned with the cultural reality of technology. Any technology that interfaces with humans requires a cultural identity. Electronic Media Arts practise is in a unique position to expose and critique the underlying cultural assumptions embedded in technological artifacts, through a technically enabled practise. This context is certainly the background of this research project that is an artistic enquiry (an enquiry emphasizing cultural aspects) of AI and cognitive conceptions of dreaming, and therefore the mind. The Context Machines, of which this research is part, are centrally concerned with autonomy and emergence, which are also central aspects of Petit Mal. Where it attempts to provide the most cultural meaning, with the least technology, this research aims to integrate cultural values and purposes into a deeply technological practise.

[79] R. P. Pérez, R. Sosa, and C. Lemaitre. A computer model for Visual-Daydreaming. In *Intelligent Narrative Technologies: Papers from the 2007 AAAI Fall Symposium, Technical Report FS-07-05*, pages 102–109, 2007.

Visual Daydreamer (V-Daydreamer) is a computational system that creates generative narratives. Narratives are an indispensable aspect cognition as they are used in the conceptualization and organization of experience. The conception of daydreaming used in the project is inspired by the "brain storming" aspect of creative writing, and is defined as "...the unrolling of a sequence of events, memories or creatively constructed images of future events..." and is not framed as a goal-directed activity.

The core of the system is the linking of events by emotional value. Researchers provide the system with a set of previous stories, sequences of events, that are annotated with emotions. The system generates "knowledge structures" through an analysis of these stories that results in "emotional contexts" being linked to one or more actions. The system is initiated with an action that results in an emotional context, which results in a selection of future actions, and so on. If an emotional context is associated with multiple actions then one is chosen randomly. Actions are manifest in predetermined animations applied to abstract shapes from Miro paintings, Australian aboriginal art and glyphs from MEXICA (the author's previous narrative system). The authors conducted an empirical study comparing narratives generated by the system with arbitrary combinations of events and narratives generated by MEXICA. Narratives generated using all three methods are visualized using V-Daydreamer. Subject reports show a very large variance in their interpretations of the three narratives, including the arbitrary selection expected to appear as non-cohesive. Most subjects classified narratives from V-Daydreamer and MEXICA as story or daydream, and were judged as having slightly higher coherence and narrative structure than the arbitrary collection of events.

V-Daydreamer defines a day-dream in humanistic terms as the brain-storming component of the creative writing process. The notion includes no reference to cognitive or neurological conceptions of day-dreaming or the construction of narratives. The system uses no artificial

intelligence or complex generative methods. The experimental results show little difference between a arbitrary collection of events and computer generated memory sequences. This paper is in strong opposition to the approach taken in this research project, where dreams are defined in cognitive/neurological terms, and will be implemented using machine learning and artificial intelligence techniques. Dreaming may simply be a special case of day-dreaming, where external stimulus provides no constraints.

[80] B. Tomlinson, E. Baumer, M. L. Yau, P. M. Alpine, L. Canales, A. Correa, B. Hornick, and A. Sharma. Dreaming of adaptive interface agents. In *CHI '07 extended abstracts on Human factors in computing systems*, CHI EA '07, pages 2007–2012, New York, NY, USA, 2007. ACM.

An adaptive interface is one that transforms in response to how it is used. For example, menu items that are often used may be reorganized to be more readily accessible, or a user model may anticipate the user's UI needs. By in large, these efforts are not successful, as the rearrangement of UI elements is frustrating and often leads to confusion and inefficient use. The metaphor of the dream is applied in order to guide the user through any interface changes that the system may initiate. This would assist users because: (1) Dreams indicate what changes may be made, and when. (2) A natural method for rejecting (interfering with dream) or accepting (allowing dream to play out) interface changes is provided. (3) Transitions between states (past and future) of the interface illustrate how the UI is changed, facilitating easier adaption.

The demonstration system is composed of a virtual world of plants, of which new types occasionally sprout. The agent is a human figure that explores this world, examining the various plants. If the user clicks on the figure, a "thought bubble" appears that reflects his mental representation of the world. It consists of an arrangement of the different types of plants, organized by "archetypes". As new plants are encountered, new mental representations are added. Occasionally, the figure falls asleep, in which case the dream appears in the thought bubble. During the dream, similar plants are reorganized into archetypes that best represent them. These archetypes are compressors of experience as many plants can be encapsulated by a single representation. During sleep, plants are decompressed and recombined during a search for an optimum representation of experience. If the user speaks into the microphone quietly the reorganization is impeded. If the user speaks loudly, the organization process is stalled as the figure wakes up. If the organization process is interrupted, the figure is not well rested and may return to sleep more quickly.

Dreams are not just representations of the past, or present, but could be imaginings of what might be in the future. Combinations of daily interface experiences could be imagined, in dream form, into future interfaces. Dreams give us the opportunity to reexamine aspects of our lives in new contexts, and therefore see them in different ways. This could encourage reflective practises as working processes are dreamt about in our interfaces. The authors propose the possibility that such a system could be combined with "MyLifeBits", an attempt to archive daily experiences. "Synthetic Dreams" could arise from combinations of daily experience "…evoking reflection on one's day, it may allow one to see connections between seemingly disparate experiences."

The conception of a dreaming process, as the compression of experience into prototypes, relates closely with other papers in this bibliography. The model of dreaming itself is quite simple, and does not make use of cognitive or neurological theories. It still provides a plausible dream mechanism for a virtual agent. Further, the proposal of a "Synthetic Dream" is very close to the types of dreams expected for "Dreaming Machine", as both would be artificial dreams resulting from real-world experience. The implementation discussed in this paper circumvents the difficulties in constructing such a system, for example the neurological and psychoanalytic characteristics of dreams, as well as the development of a meaningful conceptual structure from which a dream would result.

[81] Q. Zhang. A computational account of dreaming: Learning and memory consolidation. *Cognitive Systems Research*, 10(2):91–101, 2009.

Zhang proposes a computational model of dreaming that enables a memory consolidation function. Episodic memory is correlated with the hippocampus, while semantic knowledge is correlated with the neocortex in general. Memory consolidation is the process by which episodic memory, in the hippocampus, is "consolidated into the neocortex". This is supported by recent neural recording studies that have shown that the hippocampus "replays" recent waking patterns during sleep. Additionally, the correlation of activity between neocortex and hippocampus, and the impairment of memory during sleep, support this argument for consolidation. Zhang follows Hobson's position that dreams are the random activation of brain regions. This argument is partially based on the existence of pontine-geniculate-occipital (PGO) waves that occur during REM sleep, according to the AIM model, and the disjointed narratives of dreams. The randomness of dreams is used as an argument that they cannot have an intellectual function. Zhang does not agree, and proposes that random activation of memories can contribute to memory consolidation, giving function to randomness. Previous connectionist computational models of dreams have not shown such a connection, and tend to contradict psychoanalytic accounts of dreaming.

The system proposed in this paper is an extension of a previous work, the AI counter, which learns concepts from one dimensional stimuli. A concept is a feature that is common to all stimuli. It results when a feature of stimulus is abstracted and generalized. The learning of higher level concepts is only possible once the lower level concepts that constitute it are learned. This hierarchical learning is a key point in the system design. The system architecture is inspired by a "split-brain" hypothesis, where one hemisphere is specialized for the processing of representations, while the other is specialized for the experience of low-level features ("symbols" in Zhang's vocabulary). The design is composed of two hemispheres, the "symbol" and the "representational", where each has their own memory mechanisms, and are connected by a "bundle" corresponding to the corpus callosum. Memories are stored in interlocked loops of "single memories" that interact in order to store a pattern. Communication along the bundle allows representations on one side to be associated with low level features on the other side. The result is that representations initiate the recall of symbols, while symbols initiate the recall of representations. Concepts that are learned by the system are "zero" (no pattern), "one" (single peak in temporal window) and "tally" (multiple peaks in window). Each of these concepts is discovered in input stimulus by the system and labelled. According to the hierarchical notion of learning, "one" can only be learned once the lack of a pattern is learned. Consequently the order in which patterns are learned has a large effect on learning. If the patterns are presented in the proper order the system learns in one epoch, but learning can take considerably longer if stimuli is not presented in the correct order, as high level concepts are ignored until prerequisite concepts are learned.

The AI Dreamer is a solution to the problem of stimuli ordering, seen in the AI Counter, through the addition of an episodic memory system that corresponds to the hippocampus. Damage to the hippocampus in humans results in deficits in episodic memory and is correlated with intermediate memory storage. The episodic memory in AI Dreamer is composed of a number of memory units that each store a segment of external stimulus. These segments are structured such that the whole sequence can be recalled in sequence, or randomly. A control unit changes how episodic memories are retrieved—random or in proper sequence. When trained using the same data as AI Counter, the concepts that cannot be learned, because they are too high level, are rejected and passed onto the episodic memory where they are stored. If the training data is fed in the exact opposite order, all patterns are stored in episodic memory until the final base case, where the system learns as normal. Once the stimulus has ended, the system has only learned the base case. This is where dreaming comes in. Segments of stimulus are randomly activated in episodic memory, which are presented to the learning system as if they were external stimulus. As random patterns are presented to the learning system, the next concept in the hierarchy will eventually be recalled. The learning system learns from previously stored experience. This success proves that random memory activation can still have a functional purpose.

Zhang continues to review a number of features of dream consciousness that have already been discussed elsewhere in this bibliography. Three key features of dreaming are: dreams are random, repetitive, and include a lack of self-reflection or self-control. Zhang argues that the AI dreamer satisfies these qualities of dreaming, and therefore is more consistent with psychoanalytic evidence than previous computational models. The system obviously uses random memory activation. The hierarchical conception of learning requires that patterns are presented in a particular order. This order is only likely to occur with repetitive random activations. The system "... does not have a necessity of a central controller in regulating its learning process", and therefore lacks self-reflection and self-control.

The system described in this paper is highly relevant to this project because it is one of the few, and perhaps the only recent, rigorous computational model of dreaming. The Boolean input stimulus, used to train the system, is so simplistic that it makes supporting the argument for conceptual learning difficult to accept. Further, the work depends on the AI Counter, which is not described in sufficient detail to understand how concepts are generated in the system. A major argument for the random account of dreaming, according to Hobson, is the existence of PGO waves that occur during REM sleep. This argument does not hold up to evidence that dreams occur in NREM sleep, whereas PGO waves do not. Other considerations of dreams as random are based on experiential measures, and could be the result of the random activation of concepts, rather than the random activation of the perceptual system, as argued by Hobson. This research takes the position that dreams are activations of concepts that result in mental images. Whether these activations are random, or produced by latent activation during recent waking, is not known. Latent activation could explain the correlation between recent waking experience and dream experience. A split-brain conception of the mind has not been considered by this research. If this hypothesis is correct, then there should be evidence for increasing development of the "representation" side of the brain during the emergence of operational development in infants. The argument that Zhang's model better fits psychoanalytic evidence is problematic considering many features of dreams, for example emotional content, and lack of self-awareness, are lacking.