Voiceover Narration and Audio-Visual Imagery in Non-Fiction Film

Jeremy Lines

ORCID 0000-0002-1148-4755

Submitted in partial fulfilment of the requirements of the degree of

Master of Fine Arts (Film and Television)
(By Creative Work and Dissertation)

AUGUST 2015

School of Film and Television

Faculty Of The Victorian College Of The Arts
And The Melbourne Conservatorium Of Music

The University of Melbourne

Produced on archival quality paper

Abstract

This project investigates the relationship between audio-visual imagery and voiceover narration in non-fiction film. This thesis examines that relationship as a particular case of the broader relationship between perception and language in human cognition. I review arguments that the meanings of language are grounded in concepts acquired through perception and action, through embodied interaction with our environments. Despite this dependence of language on perception, cognitive science research shows that language exerts a significant influence over perception. For example, language has been shown to modulate visual processing at an early stage, affecting what we consciously see and remember, and attenuating bottom-up cognitive processes.

I argue that the audio-visual (AV) imagery in non-fiction films is perceptually realistic, since it addresses a subset of the same perceptual abilities we use to perceive our environments. We might therefore expect the influence of voiceover narration on our perception of AV imagery to be similar to the influence of language on perception more generally. Several film theorists, including Michel Chion and Bill Nichols, have noted such an influence in their writings.

My creative works are concerned with a number of issues raised by the philosophical and scientific study of the mind. I have experimented with the form of these video works, separating AV imagery and voiceover narration. The resulting form diverges from the most widely used structure in non-fiction films, the expository mode, in which AV imagery serves to illustrate the narrative content of the voiceover. The evidence presented in this thesis indicates that separation of these content streams will diminish the influence of language on viewers' perception of AV imagery. The resulting epistemic independence of AV imagery in my video works emphasises the act of perception as central to questions on the nature of cognition and consciousness.

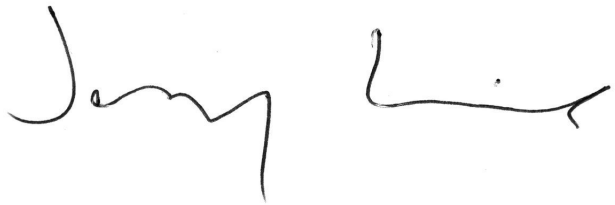
Declaration

This is to certify that:

(i) this thesis comprises only my original work towards the Masters of Fine Art degree,

(ii) due acknowledgement has been made in the text to all other material used,

(iii) the thesis is 19986 words in length, inclusive of footnotes, but exclusive of tables, maps, bibliographies and appendices.



Jeremy Lines

Acknowledgements

My thanks go to Paul Fletcher, Lecturer in Animation at the School of Film & Television at the Victorian College of the Arts, University of Melbourne, for his advice and enthusiasm throughout this project.

To my partner Rina Hadi, thank you for your encouragement and support.

I am very grateful to all those whose generosity with their time and hospitality made the video work possible: Leonie van Eyck, Salome Harris, Sebastian Harris, Glenn Norman, Trefor and Yulitta Owen, and the people of Ngukkur, particularly Andy Luckaman and Roy Natilma. Without your help, this project would not have been possible.

I would also like to express my appreciation for the financial support extended to me during my candidature via an Australian Postgraduate Award.

Abstract ..... **i**
Declaration ..... **ii**
Acknowledgements ..... **iii**
List of Illustrations **1**
Introduction **2**
Definition of Key Terms **4**
Audio-Visual (AV) Imagery 4
Cognitive Science 4
Consciousness 5
Non-Conscious 5
Perception 5
Real-World Perception 6
Symbol 6
Viewer 6
1. Embodied Cognition and The Language of Thought **7**
The Historical Identification of Thought with Symbolic Representation 7
The Trials of Computationalism: GOFAI & LOTH 8
Embodied Cognition 10
Mary The Colour Scientist 11
Conceptual Metaphor 12
The Evolution of Animal Cognition 13
Chapter Summary 14
2. Bottom-Up and Top-Down Cognition **15**
Bottom-Up and Top-Down Processing in Brains and Computers 15
Distributed Representation 17
The Scope and Significance of Bottom-up Processing in Human Cognition 19
Insight Problem-Solving 19
Perceptual Expertise 19
Multi-Factor Decision-Making 20
Unconscious Thought Theory and The Capacity Constraints of Conscious Thought 20
Heuristics and Bias in Non-Conscious Cognition 21
Why Is Bottom-Up Thought Non-Conscious? 22
Psychology and Bottom-Up Cognition 23
Hybrid Systems: The Interaction of Bottom-Up and Top-Down Cognition 24
Top-Down Processes in Perception 24
The Innocent Eye 25

The Influence of Language on Visual Perception 27
Chapter Summary 30
3. The Perceptual Realism of AV Imagery 31
Film Realism 31
Bazin's Classical Realism 32
The Language of Cinema 33
Currie's Perceptual Realism 35
Film Phenomenology and Embodied Simulation 37
Film Phenomenology 38
Mirror Neurons and Embodied Simulation 39
The Limitations of Perceptual Realism 40
The Limitations of Real-World Perception 42
4. Voiceover and Image Theory 43
The Expository Mode 43
Chion and the Dominance of Language 44
Variations on the Evidentiary Relationship in Non-Fiction Film 46
The Influence of Image Duration On Bottom-Up Perception 48
Chapter Summary 49
5. The Video Project 50
Overlaps and Correspondences 51
The Viewer as Homunculus 53
The Long Take 55
Experiments in Watching 56
Chion and the “Microstructure of the Present” 57
The Long Take in Non-Fiction Film 59
The Acousmatic Voice 59
The Fragmented Voice 61
6. Conclusion 64
Bibliography 67
Creative Works 77
Appendix: Connectionism and Artificial Neural Networks 78
A Brief Description of an Artificial Neural Network 78
Parallel Processing 79
Bottom-Up and Top-Down Interaction in Neural Networks 80
Two Consequences of Distributed Representation 80
Learning in Neural Networks 82

List of Illustrations

Figure 1: A schematic representation of an artificial neural network.

Image derived under the GNU Free Documentation License from https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg, by user Cburnett.

Introduction

I began this research project with an intention to address some of the issues raised by the scientific study of the mind. I believe many of the findings of the cognitive sciences (an umbrella term, which includes the fields of neuroscience, experimental psychology, philosophy, and artificial intelligence) are of deep and general significance. They are also often troubling, as they undermine many of our ingrained beliefs about the mind. They reveal our cognitive biases, while casting doubt on the nature of free will, the existence of the self as an entity, the central role of consciousness in our thoughts and behaviour, and the extent of our introspective access into our minds. Such issues are central to the content of the video project.

From a filmmaker's perspective, I am interested in the question of how this scientific research might inform the formal properties of a non-fiction film. As my research progressed, my interest focused on how the cognitive sciences might inform our understanding of the relationship between audio-visual (AV) imagery and voiceover narration in non-fiction films.

There has been little published research on this topic, despite cognitive science's substantial influence on film theory, evinced in the writings of David Bordwell and Kristin Thompson, Carl Platinga and others. Several other film theorists, such as Michel Chion and Bill Nichols have discussed the relationship between AV imagery and voiceover narration, but their analysis has not viewed that relationship from a cognitive perspective.

This thesis begins with a discussion of the broader relationship between perception and language in human cognition. It reviews historical perspectives on that relationship, and the evolution of those views in response to research and theories from cognitive science. Chapter One discusses the problems inherent in the historical identification of thought with language, and the growing focus on perception and action as the basis of animal cognition. Chapter Two reviews evidence that the relationship between perception and language is a key locus of a tension between two fundamentally different, but interacting modes of cognition, referred to as *bottom-up* and *top-down* processing.

The subsequent chapters of the thesis consider the relationship between AV imagery and voiceover in non-fiction films, in light of the above discussion of cognitive science research. Chapter Three establishes that this relationship in non-fiction films is a particular case of the broader relationship between perception and language. Chapter Four relates the previous discussions of cognitive science with film theorists' writings on the relationship between AV imagery and voiceover narration. Chapter Five describes my experiments with the formal relationship between AV imagery and voiceover narration in my creative work.

The research presented here suggests that the studies undertaken by cognitive science into the relationship between perception and language in cognition can deepen filmmakers' understanding of the relationship between AV imagery and voiceover in films. In turn, non-fiction films that combine AV imagery with voiceover narration are well placed to explore the broader relationship between perception and language in cognition.

Definition of Key Terms

For brevity, only terms not defined in the text are included here.

Audio-Visual (AV) Imagery

The argument presented in this thesis is applicable not only to film, but to all forms of perceptually realistic audio-visual imagery, including video and digital recordings, and computer-generated media.

I am using this term primarily to distinguish voiceover narration from the other content in non-fiction films. Thus *AV imagery* refers largely to non-narrational audio and images. Consideration of other possible content such as diagrams, on-screen text and non-diegetic music lie outside the scope of this thesis.

Cognitive Science

This thesis discusses theoretical and research findings from several fields of study, often grouped together under the umbrella term *cognitive science*. Boden identifies the key fields as "psychology, neuroscience, linguistics, philosophy, anthropology, AI (artificial intelligence), and A-Life (artificial life)" among others.[1]

Cognitive science has been a significant influence in film theory over the last quarter of a century, particularly through the work of Joseph Anderson, David Bordwell and Kristin Thompson, Carl Platinga and others.

The philosophy of mind I discuss is, for the most part, Analytic philosophy. It is sometimes known as Anglo-American philosophy, to distinguish it from the Continental philosophy of Brentano, Derrida, Heidegger, Merleau-Ponty and others. These Continental philosophers have provided influential insights into some of the arguments that will be explored here. Readers familiar with their work may find similarities in their philosophies with the conclusions reached more lately by the philosophers discussed in this thesis. However, detailed consideration of these Continental philosophers' work is beyond the scope of this thesis. The focus of both my video project and this thesis is Analytic philosophy of mind and cognitive science more generally.

[1] Boden, *Mind as Machine : A History of Cognitive Science.*, 1:xxxv.

Consciousness

In the absence of any consensus on what consciousness *is,* in functional or neurophysiological terms, the definition provided by Thomas Nagel in his paper "What Is It Like To Be A Bat?" is still considered to be the best working definition available:

> An organism has conscious mental states if and only if there is something that it is like to **be** that organism—something it is like **for** the organism.[2]

Thus consciousness is a synonym for subjective experience, for what it's like for humans, and some other animals, to think and perceive, to act and emote.

(See also Non-Conscious, below)

Non-Conscious

It is widely recognised that we are not conscious of much of what goes on in our brains. Book-length discussions include Eagleman, 2011, and Wilson, 2002.[3] The term non-conscious is used in this thesis to refer to all consciously inaccessible mental processes.

I have used the term non-conscious rather than subconscious or unconscious, to avoid the psychoanalytic implications of those words, particularly the reification of non-conscious processes as a psychological entity. There is now an extensive literature arguing for the more plausible model of multiple non-conscious systems. Various approaches to this model are discussed in Clark, 2007; Dennett, 1991; Gazzaniga and Ledoux, 1978; Minsky, 1986; and Wilson, 2002.[4]

Perception

The term *perception* is used here to refer to the prediction, acquisition and processing of sense data. Some writers use the term in a more restricted way, to mean the

---

[2] Nagel, "What Is It Like to Be a Bat?" 436

[3] Eagleman, *Incognito: The Secret Lives of Brains*; Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious.*

[4] Clark, "Soft Selves and Ecological Control"; Dennett, *Consciousness Explained*; Gazzaniga and Ledoux, *The Integrated Mind*; Minsky, *The Society of Mind*; Wilson, *Strangers to Ourselves.*

conscious categorisation or conceptualisation of acquired sense data. However, in the broad sense with which I am using the term, all animals with functioning sensory processes perceive.

Real-World Perception

I use the phrase *real-world perception* to differentiate perception of our physical environment from that of AV imagery. Our perception of our real-world environments is quite limited with respect to the quantity and accuracy of information our sensory systems can gather and process. AV imagery is still more limited, since it does not carry the full spectrum of information we are sensitive to, including taste, touch, temperature and proprioceptive information. See Chapter Three for further discussion.

Symbol

My use of the term *symbol* in this thesis is restricted to *conventional* symbols such as words, mathematical symbols and computer code. The meanings of such symbols are established by convention, they are not inherent in the symbols themselves. For example, when we learn a foreign language we must learn what the various words mean. We cannot intuit the meaning of novel symbols without context, since that meaning does not inhere in the symbols.

This thesis does not discuss artistic or psychological symbolism, such as the interpretation of dreams as a symbolic system.

Viewer

Though I use the term *viewer* throughout, this is largely because of the clumsiness of alternatives, such as 'perceiver' or 'experiencer'. It also helps distinguish between cases of real-world perception and perception of AV imagery. 'Viewer' should be taken to mean something like 'perceiver of AV imagery', as our sensory response to AV imagery is certainly not restricted to the visual mode. Evidence discussed in this thesis indicates our response involves sensory modes in the brain other than vision and hearing.

1. Embodied Cognition and The Language of Thought

Language has long been considered the medium of thought. For much of its history, Western philosophy devoted little attention to non-verbal cognition. It was only in the second half of the Twentieth Century that the shortcomings of linguistic and other symbolic systems as the media of cognition were decisively exposed, due to the problems encountered during research into artificial intelligence (AI). These problems resulted in renewed interest in the philosophy of Embodied Cognition, and in the Connectionist approach to AI.[5]

This chapter argues that perception is a gateway to a form of cognition that is not reliant on language or other symbols to develop concepts, make decisions and solve problems.

The Historical Identification of Thought with Symbolic Representation

The identification of cognition with symbolic representation (linguistic and numerical) is a central theme in the history of Western philosophy. In 1651, Hobbes argued reason was "nothing but *Reckoning*".[6] Leibniz planned to develop an alphabet of human thought, in which all knowledge could be formally expressed (*lingua characteristica*); this would be the medium of a calculus of reasoning (*calculus ratiocinator*).[7] Descartes considered the ability to express thought through language or signs as the only incontrovertible evidence of mind. He regarded non-verbal animals as no more than mindless automata.[8]

The strong emphasis on the symbolic representation of thought has continued into the present, particularly through the Computational Theory of Mind.[9] The central tenet of Computationalism is that thinking is the processing of information by the rule-based manipulation of symbolic representations.[10]

[5] See the Appendix for a discussion of Connectionism and Artificial Neural Networks.
[6] Hobbes, *Leviathan: Or, The Matter, Forme and Power of a Commonwealth Ecclesiasticall and Civil.*, 28.
[7] Peckhaus, "Leibniz's Influence on 19th Century Logic."
[8] Descartes, *Discourse on Method, Optics, Geometry, and Meteorology.*, 45–47.
[9] Computationalism is sometimes also known as Cognitivism.
[10] Horst, "The Computational Theory of Mind."

However, in the second half of the Twentieth Century, the shortcomings of the Computational model of mind began to be exposed by the failures of the early AI enterprise.

The Trials of Computationalism: GOFAI & LOTH

In the second half of the Twentieth Century, Analytic philosophy of mind and Artificial Intelligence research developed in parallel, each responding to the arguments and findings of the other. This parallel history is a useful lens for examining the limitations of recent historical ideas about the role of language and other forms of symbolic representation in cognition.

A major exponent of Computationalism is Jerry Fodor, who developed the influential Language of Thought Hypothesis (LOTH).[11] Fodor argued that a common linguistic structure underlies all human thought. This structure consists of symbols which are physically represented in the brain, and which are manipulated according to their syntactic, not semantic, properties (that is, by their shape, not their meaning). Fodor claims that this linguistic model of thought is necessary to explain some of the most powerful features of human cognition, such as productivity and systematicity.[12] (Productivity refers to the ability to generate novel mental content, and systematicity to the structured interrelatedness of mental content).

Early AI research predominantly modelled cognition as the rule-governed manipulation of a language-like symbolic system of representation. Now known as Classical, or as John Haugeland put it, "Good Ol' Fashioned AI" (GOFAI), this model of artificial intelligence enjoyed an initial wave of success with circumscribed problems in specially constructed environments.[13] But subsequent attempts to tackle more complex tasks in more realistic environments exposed the limitations of this model. Repeated failure resulted in a belated realisation that aspects of cognition which had previously received little attention, such as the recognition and tracking of objects, or discerning which aspects of the environment are relevant to

[11] Fodor, Jerry A. *The Language of Thought*.
[12] Aydede, Murat, and Brian McLaughlin. "The Language of Thought Hypothesis."
[13] Haugeland, *Artificial Intelligence: The Very Idea*., 112

the problem at hand (known as the Frame Problem), are complex and difficult problems.[14]

From the 1960s onward, Hubert Dreyfus argued against the assumptions of Computationalism and GOFAI, and predicted the symbol-based artificial intelligence research program would fail to achieve its goals. Using arguments based in the phenomenological tradition of Heidegger and Merleau-Ponty, he emphasised the role of human intuition and common sense, claiming human expertise could not be formalised as rules and symbols.[15]

According to Dreyfus, the consequence of GOFAI's symbolic representation is a world of innumerable isolated facts, related only as stipulated by the programmer and therefore inflexible with regard to context and purpose. The Computationalist model bears little or no resemblance to the world as we experience it.[16]

From the late 1970s onwards, GOFAI research did indeed encounter a number of problems. One of these problems, which goes to the heart of the Computationalist model, is the Symbol Grounding Problem — the question of how symbols (which have no intrinsic semantic content) are related to their meanings. John Searle's "Chinese Room" thought experiment illustrates the lack of intrinsic meaning in symbolic systems:

> Imagine a native English speaker (…) who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese.[17]

[14] Shanahan, Murray. "The Frame Problem."
[15] Dreyfus, *What Computers Can't Do: A Critique of Artificial Reason.*, 177.
[16] Ibid., 180.
[17] Keil and Wilson, *The MIT Encyclopedia of the Cognitive Sciences*.

This is a significant and thus far insurmountable problem for Computationalism. The Computationalist model, dependent on symbolic representation, appears unable to explain how a physical system (such as a brain), can acquire concepts and understand meanings.

The difficulties encountered by GOFAI provide clear evidence that reasoning with language or other symbolic systems, powerful though this is, cannot explain the wide spectrum of abilities, flexible response and general intelligence displayed in the behaviour of humans and other animals. The failure of GOFAI, and the implications of that failure for the Computational Theory of Mind, gave impetus to the development of the philosophy of Embodied Cognition, and the revival of Connectionism in AI research.[18]

Embodied Cognition

Embodied Cognition emphasises that the brain has evolved to guide perception and action in the world. The concepts with which we think are derived from our perceptual and practical activity. Language derives its meaning by referring to those embodied concepts.

> Grounding the symbol for 'chair', for instance, involves both the reliable detection of chairs, and also the appropriate reactions to them. These are not unrelated; 'chair' is not a concept definable in terms of a set of objective features, but denotes a certain kind of thing for sitting.... The agent must know what sitting is and be able to systematically relate that knowledge to the perceived scene, and thereby see what things (even if non-standardly) afford sitting. In the normal course of things, such knowledge is gained by mastering the skill of sitting (not to mention the related skills of walking, standing up, and moving between sitting and standing), including refining one's perceptual judgments as to what objects invite or allow these behaviors; grounding 'chair',

[18] There are a number of other approaches related to Embodied Cognition, including Situated Cognition and Enactivism.

> that is to say, involves a very specific set of physical skills and experiences.[19]

Our concepts and categories are not defined by rigid necessary and sufficient conditions, in the manner of explicit definitions, but are graded statistical patterns in which one item may be understood as more like a chair than another. These categories are not fixed, but evolve over time in a process of continuous learning.[20]

Mary The Colour Scientist

> Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on….
> What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it.[21]

Frank Jackson invented this thought experiment to argue that Physicalism is false, since the facts about the subjective experience of colour are not entailed by the physical facts.[22] I include the argument here only to illustrate the epistemic gap between language and phenomenal experience. The experience of colour cannot be described to someone who has never had such an experience, because our words for describing colour simply point to our experiences of colour.

Jackson's thought experiment supports the argument from Embodied Cognition philosophy — the meaning of language is derivative of (shared) experience. It underlines an implication of that argument: words are not vehicles for perceptual

[19] Anderson, "Embodied Cognition: A Field Guide," 102–103.
[20] This mode of "fuzzy", evolving concept formation is the product of bottom-up cognitive processes, which will be discussed in the next chapter.
[21] Jackson, "Epiphenomenal Qualia," 130.
[22] Physicalism is also known as Materialism: the theory that only matter and physical properties such as gravity exist. There are no non-physical (supernatural) substances or forces.

information in the way that pigments are vehicles for colour information.[23] The symbols of language are sparse representations, which point to, but do not convey embodied concepts. Thus, a linguistic description can never fully capture an experience, it can never fully convey what it's like to have that experience. Language (and other symbolic representations) cannot substitute for the embodied experiences to which they refer.

Conceptual Metaphor

In *Metaphors We Live By* and *Philosophy In The Flesh*, George Lakoff and Mark Johnson extend the arguments of Embodied Cognition, claiming that our ability to think about even the most abstract subjects relies on embodied experience.[24] According to these authors, we leverage our embodied experience into abstract realms through the use of cross-domain neural mapping. They call this neural mapping "Conceptual Metaphor", but assert that these "metaphors" are in fact hard-wired, pre-linguistic mapping between mental domains. We acquire Conceptual Metaphors automatically and unconsciously, and they are essential to our ability to think about abstract concepts.

An example of a Conceptual Metaphor is our understanding that Affection Is Warmth, expressed in phrases such as "a warm greeting" and "emotionally cold". Lakoff and Johnson suggest this metaphor is derived from our experiences of being physically held as infants. The two concepts of physical warmth and affection are experienced simultaneously, and are conflated in the infant's mind.

As a result of the physical connections between the two neural domains, the more abstract target domain inherits the inferential structure of the embodied source domain. Thus we can reason about abstract concepts by using the same inferences we use to reason about the associated source domain. For example, when reasoning about purposes, we may use the conceptual metaphor Purposes Are Destinations.

[23] There is a counter-argument that can be made here with regard to onomatopoeia, the rhythm and cadence of spoken language and so on. While I acknowledge that argument, I consider it limited and peripheral, firstly because onomatopoeia is rare in English. Secondly, and as a consequence of that rarity, while rhythm and cadence involve the use of the sounds of language, those sounds have for the most part only an arbitrary relationship to their meaning. The same is true of a word's shape on the printed page — while it is possible to use the physical properties of words for creative expression, the relationship between those physical properties and the meaning of the word is arbitrary.

[24] Lakoff and Johnson, *Metaphors We Live By*.; Lakoff and Johnson, *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*.

We can think about our goal as the destination of a journey, we can plan how to get there, and how to avoid obstacles. In this way, we leverage our embodied experience of moving through space as a means of reasoning about any purpose.

Lakoff and Johnson emphasise that language is not the source of conceptual metaphor. On the contrary, linguistic metaphors are only expressions of underlying conceptual metaphors.

> Metaphor is centrally a matter of thought, not just words. Metaphorical language is a reflection of metaphorical thought. Metaphorical thought, in the form of cross-domain mappings is primary; metaphorical language is secondary.[25]

The implication of Lakoff and Johnson's theory is that all of our concepts, including the most abstract, are founded on our embodied interaction with the world.

The Evolution of Animal Cognition

In evolutionary terms, language and the propositional rational thought it mediates, are recent phenomena. However, the existence of cognition in many animals is already evident in their flexible responses to their environments and their abilities in visuospatial problem-solving, communication, prediction and learning.

The fact that we are so closely genetically related to intelligent infraverbal animals such as other primates, and share so much of our cognitive abilities with them, strongly suggests that language-based cognition has evolved on a foundation of phylogenetically older, non-verbal mental capacities. Arguing against Fodor's LOTH, Paul Churchland observes:

> We appear to be the only species of cognitive creature on the planet that is capable of deploying the syntactic structures characteristic of language. If all cognition deploys them as the basic mode of doing business, why are the other terrestrial creatures so universally unable to learn any significant command of those linguistic structures? And if the basic modes of cognition in those other creatures are therefore almost certain to be

[25] Lakoff and Johnson, *Philosophy in the Flesh*., 123.

> nonlinguaformal in character, then why should we acquiesce in the delusion that human cognition – alone on the planet – is linguaformal in its basic character?[26]

We can conclude that language is not necessary for thought. It is highly likely that much (perhaps the vast majority) of human cognition is not language-like.

Chapter Summary

There has been a historical movement away from the assumption that language is essential to cognition, in part due to the failures of early AI systems that depended on symbolic representation. These failures led to a recognition that the meanings of symbol systems such as language are derivative of the concepts that emerge from our embodied interaction with the world — from perception and action. Words are sparse representations, pointers towards shared experiences. Inevitably, they contain far less information than their referents.

[26] Churchland, "Functionalism at Forty," 38.

2. Bottom-Up and Top-Down Cognition

Information processing can be characterised in terms of the direction of flow of information through a processing hierarchy, as either *bottom-up*, or *top-down*. At the bottom of the hierarchy is unstructured data (e.g. sensory data), at the top are abstractions. Perception and language are each closely aligned with one of these two modes of cognition — perception with bottom-up processing, language with top-down processing. As a result, the relationship between perception and language is a window into a fundamental dynamic in human cognition.

In this chapter I describe some of the characteristics of the bottom-up and top-down processing paradigms in biological brains and in artificial intelligence research. These two modes of cognition have significant differences. They are each better suited to different tasks and have complementary strengths and weaknesses. One cannot be substituted for the other.

The second part of this chapter discusses the experimental evidence that top-down processes can pre-empt and inhibit bottom-up cognition. Jonathan Schooler's experiments in *verbal overshadowing*, along with the work of Gary Lupyan and others, have shown that language has a significant influence on perception and bottom-up cognition. This has implications for non-fiction films which use voiceover narration, as it suggests that voiceover will tend to inhibit viewers' bottom-up perceptual processing of evidentiary AV imagery.

Bottom-Up and Top-Down Processing in Brains and Computers

The terms bottom-up and top-down are commonly used with reference to both computers and biological brains.

Bottom-up information processing systems learn to extract regularities from data, without requiring prior knowledge of those regularities — they are pattern-recognition systems. Here I will discuss two examples of systems implementing bottom-up processing — the human visual system and artificial neural networks.

In the human visual system, neurons at the earliest stages of processing extract basic, localised features from the visual field, such as edges (oriented bars), basic

colour and spatial frequency. The output from these lower level processes is fed forward to higher levels, where neurons have larger receptive field sizes and respond to more complex informational features. In the highest levels of the hierarchy (such as the Inferior Temporal cortex), single neurons respond to very complex objects, such as faces.[27]

An artificial neural network uses an analogous processing hierarchy to learn a pattern of similarities in images (of cats, for example), which enables it to reliably identify new images of cats. It achieves this without following programmed rules, instead refining its pattern recognition in response to feedback during iterative training. A more detailed description of an artificial neural network is included as an appendix.

Top-down cognition is the application of stored knowledge to data. The software with which we have everyday experience has a top-down structure. The knowledge represented in software programs is codified human knowledge, stored as symbols in tables, arrays and the like, and also stored in the rules for manipulating those symbols.

In our brains, top-down processes apply memories, existing concepts, schemas, and other stored knowledge, to predict and apply contextual structure to perceptual information. Deliberative reasoning is also top-down processing, applying learnt rules of inference to declarative (linguistically expressible) knowledge.

In *The Nature of Consciousness*, Piero Scaruffi discusses the design of an expert system which might be used by a bank to assess credit-worthiness. A top-down expert system would rely on the accumulated experience of banking experts, codified into a series of conditional if-then rules. A bottom-up expert system would be trained on the historical record of loans, until it could reliably identify patterns in loan applications resulting in approval and disapproval. Both systems would give similar results, but the approaches are very different.[28]

[27] Kreiman, Gabriel. "Biological Object Recognition."
[28] Scaruffi, Piero. "Artificial Neural Networks."

Distributed Representation

A key difference between bottom-up and top-down processing is the manner in which each mode represents information. Bottom-up systems do not require symbolic representations, nor are they programmed with explicit rules. In neural networks, a content item is represented by an overall pattern of response across the network. This is known as a *distributed representation* (see the Appendix for more information). Similar contents evoke similar representations, so the relationship between distributed representations and what they represent is not arbitrary. To a bottom-up system, a tiger is similar to a leopard because they have observable similarities which evoke similar representations. This is a fundamental difference from symbolic systems. The symbols 'tiger' and 'leopard' have no inherent similarity, so each must be externally defined as members of the category 'feline'. It can be seen then, that distributed representations have semantic properties; the similarity between the representations *constitutes* the category 'feline'. Since distributed representational systems are not dependent on external definitions, they are not subject to the fatal flaw of the Symbol Grounding Problem discussed in the previous chapter.

Distributed representation also enables neural networks to generalise, to respond appropriately to cases which do not precisely fit the characteristics of their training set. By contrast, symbolic systems are unable to respond to cases that fall outside their externally-supplied definitions.

> It is now widely acknowledged that trying to characterize ordinary notions with necessary and sufficient conditions is doomed to failure. Exceptions to almost any proposed definition are always waiting in the wings. For example, one might propose that a tiger is a large black and orange feline. But then what about albino tigers? Philosophers and cognitive psychologists have argued that categories are delimited in more flexible ways, for example via a notion of family resemblance or similarity to a prototype. Connectionist models seem especially well suited to accommodating graded notions of Category membership of this kind. Nets can learn to appreciate subtle statistical patterns that would be very hard to express as hard and fast rules.

> Connectionism promises to explain flexibility and insight found in human intelligence using methods that cannot be easily expressed in the form of exception free principles… thus avoiding the brittleness that arises from standard forms of symbolic representation.[29]

Distributed representation marks neural networks as significantly different to symbolic systems. Because of their semantic content and categorical flexibility, neural networks using distributed representation are a more plausible model of animal cognition than LOTH or the serial symbolic processing systems of traditional computers.

A further similarity between biological brains and artificial neural networks is the ability to learn through experience. Neural nets are not explicitly programmed, but instead learn by constantly modifying their structure in response to feedback. Humans also learn many skills through repeated practice, without following explicit rules. Most children rapidly learn to use natural language without explicitly learning the rules of grammar, or being able to articulate those rules.[30] Many young children learn to reproduce melodies in song, without theoretical knowledge of music. These are pattern-recognition skills.

Crucially, bottom-up systems are able to discern novel patterns. This ability to recognise novel regularities in the world is the basis of concept acquisition (as discussed earlier with reference to Embodied Cognition). In contrast, an exclusively symbolic system cannot acquire foundational concepts. This is a consequence of the conventional meaning of symbols.[31] This argument is central to the critique of the Computational Theory of Mind by proponents of Embodied Cognition such as Michael Anderson, who argues that our referential capacity is ultimately derived

[29] Garson, James. "Connectionism."
[30] To avoid confusion, I have generally avoided discussion of bottom-up cognition of symbolic content, and instead focused on bottom-up cognition of perceptual content. But it should be noted that bottom-up pattern-recognition systems can operate on any input, including symbolic inputs such as language, mathematic or scientific notation. Poetry, for example appears to have a high bottom-up component, with the discovery of novel patterns and linguistic relationships being central to that work. 'Insight' problem-solving (a paradigmatic form of bottom-up cognition characterised by a 'Eureka!' moment), occurs in solution to problems encoded in symbolic form, just as it does to problems understood in visual or other non-symbolic terms.
[31] Horst, "Symbols and Computation: A Critique of the Computational Theory of Mind."

from perceptual and practical activity.[32] The point is conceded by Computationalists, who must resort to an assumption that our foundational concepts are innate.

From a Connectionist or Embodied Cognition perspective, concepts are learnt patterns, derived by bottom-up processes from information flows, such as perception.

The Scope and Significance of Bottom-up Processing in Human Cognition

Apart from perception, bottom-up processes play a central role in other forms of cognition, including *insight problem solving*, multi-factor decision-making, perceptual expertise, and physical skill acquisition (such as learning to play a musical instrument, or to ride a bike).

Insight Problem-Solving

An insight problem solution is characterised by a "Eureka!" moment which follows an impasse. The flash of insight is characterised as "a sudden, unpredictable, and nonverbalizable solution discovery."[33] Insight problem-solving is a widely-discussed aspect of creative thought in both the arts and sciences. It is at the heart of numerous stories of discovery — August Kekulé's discovery of the structure of the benzene molecule for example, or Henri Poincaré's work on Fuchsian functions — but is also commonly experienced in more everyday situations.[34] Many of us have experienced a sudden insight after 'sleeping on' a problem.

Perceptual Expertise

In *Incognito*, David Eagleman describes the problems that experts in particular fields commonly have in verbally articulating their skills.[35] He discusses Japanese chicken-sexers, and plane-spotters employed by the British during the Second World War, as examples of people with expert perceptual skills acquired by trial-and-error learning of subtle visual clues. In each case, the experts could not articulate how

[32] Anderson, "Embodied Cognition."
[33] Sio and Ormerod, "Does Incubation Enhance Problem Solving? A Meta-Analytic Review.," 94.
[34] Gruber, Howard E. "Insight and Affect in the History of Science.," 408; Poincaré, Henri. "Mathematical Creation.," 326-328.
[35] Eagleman, *Incognito*., 57-59.

they performed their skilled acts of perceptual judgement. The only way novices could be trained was by guessing, and receiving feedback on the accuracy of each guess. This iterative feedback process has notable similarities to the training of artificial neural networks.

A study by Lewicki, Hill & Bizot found that participants in an experiment successfully learnt complex patterns, without conscious awareness of the existence of those patterns.[36] The participants were presented with a screen divided into four quadrants. A letter 'x' would appear in one of those quadrants, and the participants' task was to press one of four buttons in response. The appearances of the 'x' in the various quadrants were determined by a complex set of rules. None of the participants expressed awareness that they had been following a pattern. Yet participants' response times gradually improved — until the pattern of appearances was changed, causing the participants' performance to suddenly drop.

Multi-Factor Decision-Making

Multi-factor decisions involve large amounts of hard-to-assess information. Wilson and Schooler, and Dijksterhuis and colleagues, have conducted several experiments showing that conscious deliberation produces better results on choices with few variables, while non-conscious deliberation produces more satisfactory results for complex, multi-factor choices.[37] Further research by Bos and Dijksterhuis concluded that bottom-up, non-conscious processes are better able to process large amounts of information than are conscious processes.[38]

Unconscious Thought Theory and The Capacity Constraints of Conscious Thought

On the basis of their research into decision-making, Dijksterhuis and Nordgren developed a theory of unconscious thought (UTT).[39] They claim that since non-conscious thought operates over a massively parallel processing structure, it has far

[36] Lewicki, Hill, and Bizot, "Acquisition of Procedural Knowledge about a Pattern of Stimuli That Cannot Be Articulated."
[37] Wilson and Schooler, "Thinking Too Much"; Dijksterhuis et al., "On Making the Right Choice: The Deliberation-Without-Attention Effect."
[38] Bos and Dijksterhuis, "Unconscious Thought Works Bottom-Up and Conscious Thought Works Top-Down When Forming an Impression."
[39] Dijksterhuis and Nordgren, "A Theory of Unconscious Thought."

greater processing capacity than conscious thought.[40] Conscious thought must process information in series, because of the limitations of *working memory* — experiments have shown that working (also known as *short-term*) memory has a storage limit of only about seven separate items.[41] As a consequence, conscious thought is relatively slow, and is prone to error when confronted by large numbers of variables and long trains of thought. Because of its parallel processing structure, unconscious thought can process large amounts of information without such limitations.

Heuristics and Bias in Non-Conscious Cognition

Recent attention has been focused on the limitations of non-conscious cognition. Daniel Kahneman received a Nobel Prize for his work with Amos Tversky describing the cognitive biases that result from our non-conscious reliance on heuristics (rules of thumb). Their research showed the predictable, detrimental impact these biases have on our intuitive judgements and choices.[42] Kahneman's subsequent *Thinking, Fast and Slow* employs a two-systems approach to judgement and choice.[43] Kahneman contrasts the automatic, intuitive processes of System 1 with the much slower, but controlled deliberation of System 2.[44] System 1 is very quick and apparently effortless, but is prone to cognitive biases due to an unquestioning reliance on sometimes inaccurate heuristics.

Kahneman and Tversky's research has been highly influential, overturning the image of rational actors that had prevailed in classical economics and the social sciences. However, the other research programs discussed above have shown that not all non-conscious cognitive processes favour speed of response over accuracy. Many, such as insight problem-solving and multi-factor decision-making, are often extended processes. They have been shown to benefit from a period of incubation — what we commonly describe as 'sleeping on' a problem (see Sio and Ormerod for a

[40] See the Appendix for more information on Parallel Processing
[41] Miller, "The Magical Number Seven, plus or Minus Two: Some Limits on Our Capacity for Processing Information."
[42] Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases"; Kahneman and Tversky, "Choices, Values, and Frames."
[43] Kahneman, *Thinking, Fast and Slow*.
[44] Ibid., Chapter 2.

review).[45] In the experiments on decision-making conducted by Dijksterhuis and colleagues, such processes consistently produce better results than conscious deliberation. In a reversal of the arguments regarding the use of heuristics by non-conscious mental processes, Dijksterhuis and Nordgren claim that the capacity constraints of conscious thought necessitate a heavy reliance on schemas, and that this results in greater stereotyping.[46]

Why Is Bottom-Up Thought Non-Conscious?

The inability of Eagleman's experts and Lewicki et al.'s participants to articulate their skills suggests the possibility that human expertise may depend on a form of cognition which is not only nonverbal, but may be opaque to us — which we, as conscious observers of our own minds, cannot read.

Subtle and difficult-to-articulate discriminations underlie much of our understanding of the world. In social cognition for example, the quality of a person's physical movement, the timbre of their voice, not only helps one recognise that person (their identity), but can also give insight into their state of mind. Our appreciation of the aesthetic qualities of an image or piece of music involve a recognition of patterns within the work, and of subtle variation from those patterns.[47] We may appreciate the richness of an artistic performance, yet find it difficult to pinpoint and articulate why we find the work stimulating, particularly if we are not educated in that artistic field.

Similarly, our experiences of physical skill acquisition — of learning to ride a bike or to draw — involve fierce concentration and require extended practice, but we have little conscious insight into the mental processes controlling our physical responses.

Our introspective access into our bottom-up abilities is so limited that we are often unable to describe how we arrive at our decisions and conclusions beyond vague references to 'feelings of rightness', gut feelings, or intuition. At the same time, these feelings of knowing are often strong, accompanied by a sense of certitude, a sense that things have somehow fallen into place. We are not excluded from the results of our intuitions, but nor can we delve in and analyse them.

[45] Sio and Ormerod, "Does Incubation Enhance Problem Solving? A Meta-Analytic Review."
[46] Dijksterhuis and Nordgren, "A Theory of Unconscious Thought.," 97–98.
[47] Levitin, *This Is Your Brain on Music*, 110.

Two properties of bottom-up processing suggest that it may be *necessarily* non-conscious: the opacity of distributed representation and parallel processing.

Distributed representations are not semantically evaluable (human readable), since the patterns of connection weights which make up a representation have complex, opaque relations to the input data set. (See the Appendix for more detail)

While traditional computers are serial processors, biological brains and the artificial neural networks modelled on them, are parallel processors. However, human conscious thought is constrained in its capacity, as a result of its reliance on working memory (as described above). This means that the workings of the massively parallel processes in our brains cannot be conscious, though their outputs can be. A relatively slow serial processor, such as the human brain thinking conscious thoughts, is not capable of tracking such a vast array of simultaneously changing states in real time.

Psychology and Bottom-Up Cognition

Psychologists have used a number of closely related terms to describe difficult-to-articulate abilities, including *procedural*, *tacit* and *implicit*. These terms are each applied to knowledge, memory, learning, and skills — for example *procedural skills*, *tacit knowledge* and *implicit learning*. The terms are used to refer to a large number of mental activities, not all of which necessarily involve bottom-up processing.[48] However, many of the documented characteristics of these forms of knowledge are similar to those of bottom-up cognition: difficulty in articulation, performance does not follow explicit rules, and learning does not require effortful attentional resources comparable to explicitly conscious reasoning.[49] This suggests that bottom-up cognition plays a significant role in these abilities.

[48] While bottom-up processes are always necessarily non-conscious, it is not the case that all non-conscious mental activity involves bottom-up processing. For example, we may sometimes apply stereotypes in social situations without being conscious of doing so. Such stereotypes are top-down stored information structures.

[49] Wilson, Strangers to Ourselves., 26.

Hybrid Systems: The Interaction of Bottom-Up and Top-Down Cognition

Bottom-up and top-down cognition operate on different principles and are better suited to different types of computation. When both modes are implemented in hybrid systems, such as biological brains and recurrent neural networks, their interaction can be highly productive. This has been widely recognised since Wallas' 1926 analysis of creative problem-solving, in which each mode of cognition plays a crucial role.[50] According to Wallas, there are four stages of creative problem-solving: preparation, incubation, illumination, and verification. The preparation and verification stages are top-down processes, while the incubation and illumination stages (which would today be referred to as incubation and insight) are bottom-up processes.

Top-Down Processes in Perception

While bottom-up processes are central to human (and other animal) perception, top-down processes also play a prominent role. The physiology of the visual cortex in humans and some other animals features structures known as *backprojections*, through which top-down information is fed back through the hierarchy to modulate bottom-up processes.[51] The backprojections provide a pathway for us to apply our previous experience to incoming perceptual information, generating predictions about what we might expect to perceive, given the context. These predictions enable quick responses and significantly reduce the cognitive load of responding to the vast amounts of information we perceive, since we only need to process sufficient bottom-up information to confirm our predictions. Andy Clark describes the brain as "an engine of prediction".[52] Our everyday reliance on prediction is revealed in the surprise we feel when our predictions are in error — the jolting shock we feel when we miss a step on the stairs, for example.

Clearly, the adaptive advantages of reduced response time and reduced cognitive load will be maximised if top-down predictions are applied at the earliest stages of processing. Predictable objects can be almost instantly recognised and dismissed

[50] Wallas, *The Art of Thought*.
[51] Kreiman, "Biological Object Recognition."
[52] Clark, Edge Annual Question 2011: What Scientific Concept Would Improve Everybody's Cognitive Toolkit?

from further processing, allowing the perceiver to focus on unpredictable elements in their environment.

In vision, the earliest stages of processing occur in the primary visual cortex. Brain imaging (fMRI) studies confirm that predictable stimuli generate less activity in the primary visual cortex than do unpredictable stimuli.[53]

> Predictability reduces activity in early areas through feedback from higher level areas.[54]

A subsequent fMRI study by Kok et al. concluded that if perceptual signals of an object or event conform to our predictions, the top-down process inhibits the bottom-up process at a very early stage.

> Our results show that top-down expectations bias representations in visual cortex, demonstrating that the integration of prior information and sensory input is reflected at the earliest stages of sensory processing.[55]

While top-down processing affords significant adaptive advantages (through reduced response time and cognitive load), this comes at the cost of the partial suppression of bottom-up processes. As I have shown, bottom-up processes play a significant role in cognition, one that cannot be substituted for by top-down processing. It is this trade-off, with top-down efficiencies coming at the cost of bottom-up cognition, which I have referred to as the tension between bottom-up and top-down processing.

The Innocent Eye

Many people — artists, theorists and others — have long suspected that our habitual responses to familiar perceptual contents dulls our perception of those contents. Several of those expressing this idea appear to have thought that in childhood we enjoy a perceptual innocence, which is gradually occluded by learnt responses and habits of thought. In adulthood, we are quick to identify, categorise

[53] de-Wit, L., B. Machilsen, and T. Putzeys. "Predictive Coding and the Neural Response to Predictable Stimuli."
[54] Ibid., 8702.
[55] Kok et al., "Prior Expectations Bias Sensory Representations in Visual Cortex." 16275.

and dismiss the things we perceive, such that it may seem we barely notice them at all. In *Light Moving In Time*, William Wees traces the origin of a yearning to rediscover the "innocent eye" of childhood to John Ruskin, and notes its subsequent appearance in the writings of authors as diverse as Tolkein, Huxley, Salinger and Stan Brakhage.[56]

In *The Elements of Drawing*, John Ruskin argues that:

> The whole technical power of painting depends on our recovery of what may be called the *innocence of the eye*; that is to say, of a sort of childish perception of these flat stains of colour, merely as such, without consciousness of what they signify – as a blind man would see them if suddenly gifted with sight.[57]

Brakhage asks us to imagine "an eye which does not respond to the name of everything but which must know each object encountered in life through an adventure of perception."[58]

Paul Taberham has suggested that Ruskin's argument can retrospectively be interpreted as a wish for bottom-up perception without top-down influence.[59] But the presence of backprojections in the visual processing regions of the brain indicates that this was a forlorn hope. Top-down influence on perception is not an optional software overlay of human culture and education, it is a physical feature of the visual anatomy of humans and some other animals. There is no possibility of returning to an "innocent eye".

Taberham suggests that the skill of the painter (John Constable in his example) relies instead on *retutored* vision, which requires more schemata and "eye training" for engaging with the world, and is in this sense radically top-down.[60]

I would suggest that Taberham's may be only a partial explanation. Certainly a painter may learn from a teacher that sunlit grass (to use Ruskin's example), can be "a peculiar and somewhat dusty-looking yellow" rather than green.[61] But this is also

[56] Wees, *Light Moving in Time: Studies in the Visual Aesthetics of Avant-Garde Film*.
[57] Ruskin, *The Elements of Drawing: In Three Letters to Beginners.*, 4.
[58] Brakhage, "From *Metaphors on Vision*," 120.
[59] Taberham, "Bottom-Up Processing, Entoptic Vision and the Innocent Eye in Stan Brakhage's Work." 3.
[60] Ibid., 7.
[61] Ruskin, *The Elements of Drawing: In Three Letters to Beginners.* Endnote 1.

something the painter may discover themselves through observation — particularly if, like Constable, they paint in *plein air*, rather than from memory in their studio. In *plein air*, the mismatch between what the painter sees and the image they produce is stark, demanding that the painter looks again, more closely. The *plein air* method encourages sustained attention on perceptual sources, enabling the painter to learn new things about those sources.

So the truism that learning to paint is learning to see, may be restated in the more cognitivist terms of Andy Clark:

> Mismatches between the prediction and the received signal generate error signals that nuance the prediction or (in more extreme cases) drive learning and plasticity.[62]

A brain-imaging study by the previously-cited Kok and colleagues showed that attention reverses the attenuation of bottom-up processes.[63] This indicates that while top-down processes pre-empt and inhibit certain bottom-up functions at an early stage of perceptual processing, observation beyond that initial prediction and recognition phase will provide new data and renewed stimulus to our bottom-up cognitive abilities.

The Influence of Language on Visual Perception

Current research indicates that language has a significant top-down influence on vision, extending to very low levels in the visual hierarchy. Using eye-tracking hardware to study the interaction between vision and language, studies have shown that listening to speech directs visual attention. Indeed, the linkage between listening to linguistic expressions and eye movements to related objects is so tight it is considered "a fundamental property of the comprehension system".[64]

Numerous other studies have found linguistic effects on colour perception, motion perception, visual search and categorisation. (For an overview, see Lupyan, 2012).[65]

---

[62] Clark, Edge Annual Question 2011: What Scientific Concept Would Improve Everybody's Cognitive Toolkit?
[63] Kok et al., "Attention Reverses the Effect of Prediction in Silencing Sensory Signals."
[64] Andersson, Ferreira, and Henderson, "I See What You're Saying," 210.
[65] Lupyan, Gary. "Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis," 1.

Gary Lupyan and colleagues have performed many experiments on the effect of language on visual perception. Their experiments have shown that simply speaking the name of an object while performing a visual search makes it easier to find.[66] In another experiment, Lupyan and colleagues found that applying linguistic labels to novel, subtly-different visual elements improved participants' response times for those elements.[67] Hearing the name of a briefly-shown object enhanced participants' detection sensitivity for that object.[68] Each of these studies show a linguistic effect on early visual processing.

Lupyan and Ward demonstrated that hearing a verbal cue can boost an invisible object into awareness. The researchers used continuous flash suppression to suppress participants' visual awareness of images of familiar objects. They found that hearing the name of an object that would otherwise not have been seen "boosted it" into conscious perception. Hearing an incorrect name did not have the same effect. The authors concluded that "a weak bottom-up signal, combined with the top-down signal produced by the label may be sufficient to propel the percept into awareness."[69]

Lupyan's experiments also provide evidence confirming that the top-down effect of language attenuates bottom-up perceptual processes. In a series of experiments in 2008, Lupyan asked participants to simply categorise a number of images (for example, categorising chairs and tables as 'chair' or 'table').[70] He found that the participants had poorer subsequent recognition of those items that they had labelled, compared to control images. Those items that were highly typical of their categories were remembered even less well. Lupyan concluded that the top-down effect of deliberately categorising images modulates bottom-up perceptual processes. He writes the "transient effects of labels on perception described above may be special cases of normally occurring top-down modulations of vision by linguistic, contextual and other 'cognitive' factors".[71] He suggests that the mechanism by which language modulates bottom-up perceptual processing is by initiating a top-down predictive processing response.

[66] Lupyan and Swingley, "Self-Directed Speech Affects Visual Search Performance."
[67] Lupyan, Rakison, and McClelland, "Language Is Not Just for Talking: Redundant Labels Facilitate Learning of Novel Categories."
[68] Lupyan and Spivey, "Making the Invisible Visible."
[69] Lupyan and Ward, "Language Can Boost Otherwise Unseen Objects into Visual Awareness," 14199.
[70] Lupyan, "From Chair to 'Chair.'"
[71] Lupyan, "Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis.," 6.

> One possibility is that processing an object name initiates a volley of feedback activity to object-selective regions of cortex such as IT (Logothetis and Sheinberg, 1996), producing a predictive signal or "head start" to the visual system.[72]

This indicates that language is deeply integrated into our phylogenetically older top-down predictive and identificatory processing. As discussed, these top-down processes operate at a very early stage of perceptual processing, before bottom-up processes complete. Thus, language has the ability to pre-empt and modulate bottom-up perceptual processing.

Lupyan and colleagues' categorisation studies add weight to research done by Jonathan Schooler and others in the 1990s. Schooler and Engstler-Schooler found that memory for a previously seen face decreases if participants verbally describe the observed face before attempting to identify it among a group of alternatives.[73] Schooler described the effect as *verbal overshadowing*. Later studies found it generalised not only to other perceptual modalities (including music, voices and wine), but also to other difficult-to-articulate cognitive activities, such as insight problem solving and affective decision-making.[74]

Schooler concluded that non-verbal cognition is negatively affected by verbalisation, and might occur "any time individuals attempt to verbalize cognitions that involve inherently non-verbal performance".[75] He hypothesised that "verbalization produces a 'transfer inappropriate processing shift' whereby the cognitive operations engaged during verbalization dampen the activation of brain regions associated with critical non-verbal operations."[76]

For a period, Schooler's work was cast into some doubt due to a *decline effect* — follow-up studies by Schooler and others found diminishing evidence for the phenomenon. However, subsequent studies such as Brown and Lloyd-Jones (2003), Wickham and Swift (2006) as well as Lupyan's 2008 research have also found

[72] Lupyan, "Linguistically Modulated Perception and Cognition, 7.
[73] Schooler and Engstler-Schooler, "Verbal Overshadowing of Visual Memories."
[74] Schooler, Ohlsson, and Brooks, "Thoughts beyond Words."
[75] Schooler, "Verbalization Produces a Transfer Inappropriate Processing Shift," 993.
[76] Ibid., 989.

evidence of a verbal overshadowing effect.[77] In a recent large-scale study, thirty-one laboratories each independently replicated Schooler and Engstler-Schooler's experiment. The meta-analysis of all the experiments found strong support for the verbal overshadowing effect.[78]

Chapter Summary

There is strong experimental evidence from a number of sources corroborating the contention that while language is semantically derivative of bottom-up processing, it also has an ability to inhibit or modulate bottom-up processing. It achieves this by being tightly integrated into phylogenetically older top-down processes, such as predictive processing. This powerful influence of language on perception has significant implications for the relationship between AV imagery and voiceover narration in non-fiction films. Those implications will be discussed in the following chapters.

[77] Brown, Charity, and Toby J. Lloyd-Jones, "Verbal Overshadowing of Multiple Face and Car Recognition."; Wickham, Lee H. V., and Hayley Swift, "Articulatory Suppression Attenuates the Verbal Overshadowing Effect."
[78] Alogna et al., "Registered Replication Report: Schooler and Engstler-Schooler (1990)."

3. The Perceptual Realism of AV Imagery

The previous chapters have argued that while language is derivative *of* perception, it has a significant influence *on* perception. Experimental evidence indicates that the likely mechanism is the integration of language into phylogenetically older top-down mental processing. Top-down processes, such as predictive processing, achieve response efficiencies by inhibiting bottom-up processing. However, the corollary of this increased efficiency is a reduction in inherently bottom-up processes, such as embodied concept acquisition, pattern-recognition and insight problem-resolution. I have referred to this as the tension between bottom-up and top-down cognition.

This chapter explores the parallel relationships between real-world perception and language on the one hand, and between AV imagery and voiceover narration in non-fiction films. I argue that AV imagery is *perceptually realistic*; that perceiving AV imagery uses a subset of the same perceptual abilities as does real-world perception. The relationship between AV imagery and voiceover narration in non-fiction films can therefore be considered a particular case of the broader relationship between perception and language in human cognition. On the basis of the evidence reviewed in the previous chapter regarding the influence of language on perception, we can infer that voiceover narration can modulate viewers' perception, and thereby inhibit their bottom-up cognition of AV imagery.

To make the argument that AV imagery is perceptually realistic, I will briefly review the historical discussions around film realism, including Classical Realism and its critics, and Currie's arguments for Perceptual Realism. I also consider two recent theories (Film Phenomenology and Embodied Simulation), which claim that our perception of AV imagery involves sensory modalities other than hearing and vision. While scientific evidence supports those claims, I note the perceptual gap that remains between perception of AV imagery and real-world perception.

Film Realism

Audio-visual media seem able to reproduce the visual and aural properties of the world much as we perceive them during real-world perception. Because of this realism, we might expect that the relationship between AV imagery and voiceover

narration in non-fiction films may have strong parallels with the broader relationship between real-world perception and language in general. This would suggest that audio-visual media are well placed to explore that broader relationship.

There seems no strong reason to doubt that we perceive and comprehend voiceover narration in much the same way as we perceive and comprehend the spoken word more generally. But the same cannot be said of AV imagery. Beginning with Structuralism, many film theorists have argued that André Bazin and other proponents of Classical Realism were naive, and that film imagery is primarily a system of signs which has much in common with language, and is read by viewers as a text. If this is the case, audio-visual media may be no better placed than books with regard to an exploration of the relationship between language and perception.

I argue below that Gregory Currie's arguments for the perceptual realism of AV media are more persuasive than those discursive theories of film. While not denying that images can carry a sub-textual or cultural meaning, I argue that AV imagery is not primarily symbolic, but addresses a subset of the same perceptual abilities in the viewer as does perception of the real world.

Bazin's Classical Realism

André Bazin claimed that photographs, film and other recorded media are *essentially* realistic due to an indexical or transparency relationship between the image and its referent. A photograph has an identity relation to the object photographed, because of the mechanical causal process of photography. Bazin likened photography to a sculptural mould, automatically tracing an object.

At times, Bazin also suggested a closer ontological identity between the photographed image and its subject, claiming "we are forced to accept as real the existence of the object reproduced, actually re-presented", and "The photographic image is the object itself".[79]

There has been criticism of Bazin's Classical Realism from a number of sources, some of which I will discuss below. From today's perspective, with ongoing rapid advances in computer generated imagery (CGI), it is increasingly clear that an

[79] Bazin, *What Is Cinema?*, 1:13–14.

indexical relationship is no longer essential for realism. It is probable that in the near future, synthetic images will be perceptually indistinguishable from recorded images. That point of indistinguishability has already been reached for images of manufactured objects (clocks, cars or laptop computers, for example), and appears to be inexorably approaching for images of organic substances and biological beings.

As a result, we can no longer be certain of the indexical status of the objects depicted in films. Indexicality is not an essential property of a realistic image. If an indexical relationship exists between an image and what it depicts, that indexicality is part of the history of the image's making, but is not discernible in the image itself. Indexicality is therefore dependent on the knowledge, judgement or belief of the observer. Such knowledge and belief can be an extremely important factor in the experience of viewing AV images, but a judgement regarding the existence of an indexical relationship based on observable properties alone is becoming less tenable.

My argument does not rest on an indexical relationship between AV imagery and subject. In order to consider the relationship between voiceover narration and AV imagery as a particular case of the broader relationship between language and perception, AV imagery need only be perceptually realistic. Following Currie, I will argue that AV imagery is perceptually realistic, since it addresses a subset of the same perceptual abilities in the viewer as does real-world perception.

The Language of Cinema

Some of the most trenchant critics of Bazin's realism were theorists influenced by Ferdinand de Saussure's structural linguistics. The Structuralists and subsequent discursive (semiotic, Marxist, psychoanalytic) theorists rejected Classical Realism as naive, arguing that filmed images are the coded product of an ideology, not a window onto the world.

I certainly do not dispute the existence of an ideological dimension of the cinema. Enculturated responses to films may take the form of implicit assumptions, inferences and schemata. These operate as learnt, often unconscious, top-down influences on cognition. Cinema, like other media, can inculcate, reinforce or otherwise modify those enculturated responses, and so has an ideological

dimension. However, the foundation of the above-mentioned discursive theories in Saussurean linguistics limits their ability to address the aspects of AV imagery and perception that are the focus of my project.

De Saussure's linguistic theory has similarities with Jerry Fodor's Language of Thought Hypothesis (LOTH), and shares some of its shortcomings. De Saussure proposed a dyadic model, in which the sign is composed of two elements: a *signifier* (a perceived sound image or visual image), and a *signified* (an immaterial concept).[80] Languages are systems of signs in which the relationship between the signified and signifier is arbitrary, determined by convention.[81] Different signifiers, such as sounds in different languages, can refer to the same signified. *Anjing hitam* has the same meaning as 'black dog'.

Like Fodor's Language of Thought, Saussurian linguistics is a formal, abstract system: in semiology, signs receive definition not by reference, but through their differential relation to each other. So the meaning of the word 'dog' is defined not by reference to a species of animal, but by its relation to, and difference from, other words (cat, barking, puppy, and so on). This relativism means Saussurian linguistics is subject to the same Symbol Grounding Problem which critically weakens LOTH.

The linguistic model now seems to have been an unfortunate choice to apply to AV media, since it has difficulty in addressing the non-symbolic attributes of those media. Christian Metz recognised some of the limitations of the linguistic model, noting that while linguistic denotation is largely conventional, filmed images usually resemble the subject they denote.[82] A person who has previously encountered black dogs must still learn the conventions of the Indonesian language to understand the phrase *anjing hitam*, but that person does not need to learn any conventions to recognise a filmed image of a black dog. Acknowledging this deficiency of the linguistic model with respect to recorded images, subsequent theorists of cinema semiotics such as Peter Wollen and Stephen Prince adopted C.S. Peirce's tripartite theory of signs. Peirce's semiotics is a general theory of signs, not restricted to language. He proposed three elements of a sign: *icons* (pictorial, formal resemblance), *indices* (relation in fact) and *symbols* (arbitrary correspondence). The

[80] de Saussure, *Course in General Linguistics.*, 76.
[81] Ibid., 78.
[82] Metz, *Film Language: A Semiotics of the Cinema.*, 108–9.

iconic and indexical elements make Peirce's system more appropriate for film analysis.

Yet the influence of the linguistic model remains pervasive in contemporary discussions of film. We often speak of film as a 'text' to be 'read', and refer to 'film grammar' and 'film language'. Film theory very often operates as a form of hermeneutics.

> The debt to Saussurean linguistics is apparent in the axioms that the connections between cinematic representation and the world are, in all important respects, a matter of historical or cultural coding and convention, that is, that filmic representation is a matter of symbolic rather than iconic coding and that a viewer, rather than perceiving a film, "reads" it.[83]

Currie's Perceptual Realism

In *Image and Mind*, Gregory Currie argues for the non-conventionality of images.[84] The core of Currie's argument is that natural languages are both productive and conventional, and that this implies certain other entailments.

Languages are productive in that an unlimited number of sentences can be uttered and comprehended. It is possible to utter and to comprehend sentences that no one has ever heard before. Languages are also conventional and so their meanings must be learned. Productivity precludes language being learned sentence by sentence. So the meanings of sentences must be specified recursively; we understand the meaning of a sentence because we understand the meanings of the words it is composed of, and the rules of composition. This entails that language is molecular: its sentences are built out of independently meaningful units (i.e. words), which can be characterised as "atoms of meaning".[85]

Advocates of a language of cinema claim that cinematic images are language-like because they are both productive and conventional. Currie agrees that cinematic images are productive; there are an unlimited number of scenes that can be

[83] Prince, "The Discourse of Pictures," 18.
[84] Currie, Gregory. *Image and Mind: Film, Philosophy and Cognitive Science*.
[85] Ibid., 122.

represented, and viewers have no trouble grasping what filmed images represent even if they have never seen them before. However, Currie argues that images are not atomic, since "every temporal and spatial part of the image is meaningful down to the limits of visual discriminability."[86] Images do not contain the functional equivalent of words in sentences.

If images are productive, but have no atoms of meaning and so are non-recursive, they cannot be conventional. The proponents of cinematic language, according to Currie, fail to distinguish between what is conventional in the manner of linguistic systems, and what is merely influenced by convention. Certainly there are conventions in cinematic images, such as the styles of shot composition characteristic of film noir, but "this is not grounds for saying the meaning of the image is *itself* conventional in the sense that meaning in natural language is".[87]

Instead of conventionality, pictures have natural meaning. That is to say, "we understand what is depicted on screen by employing our visual capacities to recognise the entities depicted."[88]

Currie argues that pictures represent by *likeness* (the resemblance of images to what they represent). We recognise what an image depicts by its spatial features, including shape, colour and so on. These spatial features trigger our capacity to recognise the depicted object, in much the same way as the depicted object would itself trigger those recognition capacities. This means images are realistic, to varying degrees.

Currie gives the example of a picture of a horse, which triggers his brain's horse-recognition capacity. Currie places his horse-recognition capacity in a low, sub-personal stratum of the brain's hierarchy. It is a quick-and-dirty horse-recognition capacity, which is "prone to be fooled by donkeys at dusk, stuffed horses and, in particular, pictures of horses."[89]. It is up to higher-level capacities, which integrate information from other senses, along with contextual information, to make the judgement that the picture of a horse is not actually a real horse.

[86] Currie, *Image and Mind*, 130.
[87] Ibid., 131.
[88] Ibid., 130–131.
[89] Ibid., 85.

We do not need to commit to this strictly hierarchical, somewhat bureaucratic model to accept the thesis of perceptual realism. Rather than a low-level functionary who is fooled and then overridden by the "higher-ups", we can posit a recurrent perceptual network with evolving interactions between bottom-up and top-down architectures. Bottom-up processes have the ability to recognise the isomorphic similarities of a given horse image to previous instances of horses (or horse images) the viewer may have encountered in the past. Top-down processes can accelerate this process, as well as integrate contextual information.

Currie's thesis of perceptual realism is a plausible and coherent account of image recognition. We recognise and understand the depictions of recorded audio-visual imagery using the same perceptual abilities we use to recognise and understand objects and events in the real world. Our recognition of imagery does not rely on a language-like conventional relationship between an image and what it depicts. Nor does the realism of AV imagery rely on an ontological or indexical relationship to its source. Instead, it relies on a similarity between the way we perceive and understand AV imagery with the way we perceive and understand our environments.

There are limitations to perceptual realism, since AV imagery addresses only a subset of our perceptual abilities. I address some of these limitations in the next section.

Film Phenomenology and Embodied Simulation

Two recent theories, Film Phenomenology and Embodied Simulation, each claim that perception of AV imagery recruits sensory-motor areas outside the specialised visual and auditory processing regions of the brain. Their emphasis on cross-modal, embodied perception involves a shift away from discursive theories towards the model suggested by Embodied Cognition. Their claims of cross-modal perception may extend the perceptual realism of AV imagery beyond the modalities of vision and hearing.

The Embodied Cognition theorists discussed earlier emphasise that our understanding of the world is grounded in our experience of physical interaction with it. Rather than a cognitive model in which the brain activates discrete symbolic

concepts in response to data from highly modular senses, our understanding of our environments involves the non-symbolic integration of all aspects of our embodied interaction.

The film theories discussed below argue that the perception of modally-restricted media such as film can activate these wider networks of embodied understanding.

Film Phenomenology

Vivian Sobchack claims that our experience of film is not limited to the perceptual modes of sight and hearing.[90] Sobchack argues that our perceptual modes are not as distinct or modular as traditionally thought. She emphasises a whole-of-body experience of film, in which other sensory modes such as touch and taste are activated as we 'watch' a film.

> We do not experience any movie only through our eyes. We see and comprehend and feel films with our entire bodily being, informed by the full history and carnal knowledge of our acculturated sensorium.[91]

Sobchack's stance is rooted in the phenomenological theories of Merleau-Ponty, who argued that the senses, though "discrete" (modular), communicate with each other.

Sobchack does not cite a great deal of neuroscientific research, but there is now increasing evidence supporting these phenomenological observations. Neuroscientists specialising in perception now consider it a fundamentally multisensory phenomenon (see Calvert et al., 2004, for a discussion)[92] There is evidence of extensive low-level connections between sensory modalities.[93] There is also evidence of the existence of multisensory neurons, and studies showing that sensory cortices respond to stimuli that are not of their principal modality.[94]

[90] Sobchack, *What My Fingers Knew: The Cinesthetic Subject, or Vision in the Flesh*.
[91] Ibid., 63.
[92] Calvert, Gemma, Charles Spence, and Barry E. Stein. *The Handbook of Multisensory Processes*.
[93] Fogassi and Gallese, "Action as a Binding Key to Multisensory Integration," 425; Gallese and Lakoff, "The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge," 459.
[94] Wallace, Ramachandran, and Stein, "A Revised View of Sensory Cortical Parcellation," 2167; Kayser et al., "Visual Enhancement of the Information Representation in Auditory Cortex," 19.

Crucially for Sobchack's arguments, the primary somatosensory (touch) cortex has been reported to respond to behaviourally-associated visual and auditory stimuli.[95]

Mirror Neurons and Embodied Simulation

Discovered late last century, *mirror neurons* are claimed to have important implications for the study of how our brains make sense of AV imagery.[96] Mirror neurons are located in the motor regions of the brain. They are active when we plan and perform actions — but they are also active when we observe other people's actions. This suggests that we understand observed actions by simulating them, using the brain's motor regions.

Vittorio Gallese (one of the discoverers of mirror neurons) has worked with Michele Guerra to develop the implications of mirror neurons and related mechanisms for the study of film. Their Embodied Simulation (ES) theory claims that our sensory-motor systems are not only involved in our understanding of others' actions, but also our understanding of others' sensations and emotions, the perception of camera movements and the affordances of objects shown on screen.[97]

> ES generates The Feeling of the Body [which] consists of the activation within the observer of non-linguistic "representations" of the body-states associated with the observed actions, emotions, and sensations, as if he or she were performing a similar action or experiencing a similar emotion or sensation.[98]

Recent brain-imaging (EEG) experiments on viewers of video clips have shown that viewers' mirror mechanism responses increase when the camera moves towards its subject, either by tracking with a dolly, or using a stabiliser such as Steadicam. The authors conclude that simulating the vision of a walking human observer approaching a scene increases viewers' sense of involvement by eliciting this stronger mirror mechanism response.[99]

---

[95] Zhou and Fuster, "Visuo-Tactile Cross-Modal Associations in Cortical Somatosensory Cells," 9777.
[96] Gallese and Guerra, "Embodying Movies: Embodied Simulation and Film Studies," 184.
[97] Gallese and Guerra, "Embodying Movies: Embodied Simulation and Film Studies."
[98] Ibid., 193.
[99] Heimann et al., "Moving Mirrors."

This indicates that when watching recorded images, we identify with the camera as our own point of view, and understand its movements using our sensorimotor abilities.

> We do not empathize just with the characters, but with a meta-character like the camera, whose behavior is generally "goal-oriented" and "action-packed", that is, suitable to be interpreted by our brain-body system.[100]

Like Film Phenomenology, the theory of Embodied Simulation focuses on our response to perceptual information at a sub-symbolic level, the level of embodied cognition. It supports the argument that our understanding of perceptually realistic imagery is different in kind to our understanding of language, using a non-symbolic cognitive architecture. The weight of evidence appears to be moving decisively away from the model of perception in which its fundamental mode of operation is the activation of symbolic concepts by sensory information, and the processing of those symbols using a quasi-linguistic syntax, in a language of thought.

The Limitations of Perceptual Realism

The theories of Film Phenomenology and Embodied Simulation extend perceptual realism beyond strictly audio-visual domains into brain structures that were not previously thought to be involved in the processing of AV imagery. They suggest that our perception of AV imagery activates the networks of embodied understanding with which we understand and interact with the world.

However, it is also clear that this realism is limited. There is a *perceptual gap* between AV imagery and the real world. Though seeing and hearing films may involve brain areas specialised for other sensory modes, this does not compensate for a lack of direct sensation in those other modes. The viewer's cross-modal response to an image of someone eating a mango remains a pale reflection of the real-world experience of tasting the fruit.[101]

---

[100] Guerra, "Film Style: A Motor Approach."

[101] In Chapter One, I made a related argument regarding the informational poverty, or "sparseness", of language (see Ch. 1 Mary The Colour Scientist). There is however a basic difference between linguistic symbols and perceptually realistic representations: symbols contain none of the perceptual information of

Sobchack acknowledges that cross-modal activation cannot bridge the perceptual gap:

> However hard I may hold my breath or grasp my theater seat, I don't have precisely the same wild ride watching *Speed* that I would were I actually on that runaway bus. I also don't taste or smell or digest those luscious dishes in *Like Water for Chocolate* … in the same way I would if, unmediated by cinema, they were set on the table before me.[102]

A perceptual gap also applies to Embodied Simulation. Recall the EEG experiments on camera movement in which viewers were found to be more involved in recorded scenes in which the camera approached the scene to simulate the vision of a walking human observer. The viewer perceives visual cues, which as the researchers found, increases their involvement in the scene. However, there is a lack of proprioceptive feedback in this research scenario, and I would suggest that this lack causes a significant perceptual gap for the viewer, since proprioception plays an important role in perception. Most of the changes in our perception are caused by our own motor actions. As a result, our brains have highly-developed predictive and feedback mechanisms which integrate perceptual information with motor activity. An example of this is *corollary discharge*: information from the motor areas controlling eye and body movement is used to predict and offset the changes in visual input caused by saccades and other motor activity, enabling us to see a stable world. The absence of such feedback explains the sometimes jarring effects of gyrating handheld and body-mounted camera footage. In my opinion, this lack of proprioceptive feedback means that handheld camerawork, despite signifying documentary immediacy, is much less perceptually realistic than static or smoothly moving footage. I will come back to this, with regard to my choice of shooting style, in Chapter Five. Experiments have shown that even a rat on a multi-directional treadmill in a virtual reality environment — a laboratory model which ensures that the visual world shifts in exact correlation with the animal's every move — still has an impaired understanding of space, due to the lack of motion feedback from the

---

their referent, whereas AV imagery contains a great deal (though not all) of that information. This is not a difference of degree, but a fundamental difference between AV imagery and language.

[102] Sobchack, *What My Fingers Knew*, 72.

vestibular apparatus in the inner ear.[103] Even though the rat is walking in an apparently-changing environment, at this vestibular level it knows it isn't moving through space. There is a perceptual gap for the rat.

The Limitations of Real-World Perception

Though we are generally unaware of the limitations of our real-world perception, experimental evidence clearly shows that we do not have direct, comprehensive and accurate perception of our environments. Realism is never absolute, it is always a matter of degree.

It is interesting to note how film technology has developed to exploit the limitations of the human perceptual system. Twenty-four frames per second is sufficient to overwhelm our ability to discriminate individual frames, though that frame rate would be inadequate if our visual systems had the speed of many other animals' (an insectivorous bird's, for example). Similarly, audio and video compression systems are designed to reflect the sensitivities and limitations of our perceptual systems. Many common image compression formats use colour subsampling, a technique which involves a significant reduction in chrominance (hue and saturation) information, while preserving luminance information, which our eyes are far more sensitive to.

Audio-visual images provide a limited spectrum of the information a healthy human perceptual system can gather and process. But they are realistic nonetheless, since they address a subset of the same perceptual abilities we use to perceive the real world.[104] We can conclude that AV imagery stimulates a subset of the same bottom-up cognitive processes that are involved in real-world perception. As a result, we can consider the relationship between voiceover narration and imagery in non-fiction films to be similar to the broader relationship between language and perception studied by Lupyan, Schooler and others.

[103] Ravassard et al., "Multisensory Control of Hippocampal Spatiotemporal Selectivity."

[104] As the philosophy of Embodied Cognition makes clear, real-world perception is closely integrated with action. The absence of pathways to action marks AV Imagery as quite different from most (though not all) cases of real-world perception. But this does not detract from the argument being made here: since AV Imagery addresses a subset of the same perceptual abilities as we use to perceive the real world, it is perceptually realistic.

4. Voiceover and Image Theory

This chapter reviews some of the literature that bears on the relationship between voiceover and AV imagery in non-fiction films. It also discusses styles of filmmaking which seek to modify the most commonly-used relationship (the *expository mode*), by developing contradictory or ironic relationships between AV imagery and voiceover narration.

The Expository Mode

The use of voiceover narration as the primary means of structural and conceptual organisation of non-fiction films dates almost from the inception of film sound. A well-known early example is *Night Mail* from 1936, which features a poem written by W.H. Auden.[105] Filmmakers quickly recognised that voiceover narration is an effective means of conveying concise exposition and rhetorical argument. It was rapidly adopted as a replacement for the intertitles used by silent documentaries such as *Nanook of the North*, and it continues to be a central pillar of non-fiction filmmaking.[106]

Bill Nichols identifies the *expository mode* as the most common structural mode of documentary film and television, both historically and in contemporary production.[107] The expository mode uses voiceover narration to address the viewer directly. The voiceover dominates the film's images, which are selected to support the spoken narrative.

> Expository documentaries rely heavily on an informing logic carried by the spoken word. In a reversal of the traditional emphasis in film, images serve a supporting role. They illustrate, illuminate, evoke, or act in counterpoint to what is said.… The commentary, in fact, represents the film's perspective. We take our cue from the commentary and understand the images as evidence or illustration for what is said.[108]

[105] Watt and Wright, *Night Mail*.
[106] Flaherty, *Nanook of the North*.
[107] Nichols, *Introduction to Documentary*, 171.
[108] Ibid., 168–9.

In my experience of non-fiction video production, the voiceover narration is almost always written, and often recorded before the picture edit gets beyond an initial assembly. In many cases, the narration is drafted to an advanced stage before the shoot begins. The shooting script is then developed to support the narration. The voiceover is the vehicle for the conceptual content of the program, and largely determines what will be shot. This is not always the case; projects which follow unfolding events, and projects which feature interviews, will often be scripted later in the production process, in response to the incoming information. But even in these latter cases, where segments of those programs use voiceover narration, that narration will strongly influence the selection and duration of shots.

Nichols describes this as *evidentiary editing*: the images serve to provide evidence for, or to illustrate the content of the voiceover narration.[109] Evidentiary editing is commonly found in non-fiction films whenever voiceover narration is employed, including in those programs that might be grouped under other modes in Nichols' analysis of documentary form, such as the *participatory* and *performative* modes.

Serge Daney notes that since images serve only as illustrations for the spoken word, they have a severely constrained epistemic role.

> [The voice] enters into an alliance or contract [with the viewer] that ignores the image. Because the image serves only as the pretext for the wedding of commentary and viewer, the image is left in an enigmatic state of abandonment, of frantic disinheritance…[110]

Chion and the Dominance of Language

In *Audio-Vision: Sound on Screen*, Michel Chion offers an amusing anecdote to illustrate the influence of language on our visual perception of documentary films. Chion argues that verbal statements, such as those in voiceover narration, "guide and structure our vision".[111]

[109] Ibid., 169.
[110] Daney, "Back to Voice," 19.
[111] Chion, *Audio-Vision: Sound on Screen*, 7.

It is worth quoting at length:

> An eloquent example that I often draw on in my classes … is a TV broadcast from 1984, a transmission of an air show in England, anchored from a French studio for French audiences by our own Léon Zitrone. Visibly thrown by these images coming to him on the wire with no explanation and in no special order, the valiant anchor nevertheless does his job as well as he can. At a certain point, he affirms, "Here are three small airplanes," as we see an image with, yes, three little airplanes against a blue sky, and the outrageous redundancy never fails to provoke laughter.
> Zitrone could just as well have said, "The weather is magnificent today," and that's what we would have seen in the image, where there are in fact no clouds. Or: "The first two planes are ahead of the third," and then everyone would have seen that. Or else: "Where did the fourth plane go?" — and the fourth airplane's absence, this plane hopping out of Zitrone's hat by the sheer power of the Word, would have jumped to our eyes. In short, the anchor could have made fifty other "redundant" comments; but their redundancy is illusory, since in each case these statements would have guided and structured our vision so that we would have seen them "naturally" in the image.[112]

Chion concludes that "if the film or TV image seems to 'speak' for itself, it is actually a ventriloquist's speech."[113]

There is a striking parallel here to the experimental evidence amassed by Richard Andersson, Gary Lupyan and others, discussed in Chapter Two. That evidence suggests that the dominance of language will cause us to attend to exactly those elements mentioned by the voiceover, neglecting many other observable details. Jonathan Schooler's experiments also suggest that verbal description will tend to inhibit our processing of images, reducing our memory of them and making it less

[112] Chion, *Audio-Vision*, 6–7.
[113] Ibid.

likely that they will be factors in our bottom-up cognitive processes. These psychological experiments are supported by the fMRI studies cited previously.[114]

To the extent we are told what we are looking at, the less we may see.

Variations on the Evidentiary Relationship in Non-Fiction Film

Nichols points out that imagery need not directly illustrate voiceover narration; there are alternatives to evidentiary editing. Filmmakers can also develop a contrapuntal or ironic relationship between these two streams of content. The dialectic between the two may then yield a more complex meaning than the more straightforward structure of voiceover narration with evidentiary editing.

Nichols discusses two documentaries that construct conflicts between interview testimony and the AV imagery that accompanies them.

The Life and Times of Rosie the Riveter features interviews with five women, former factory workers during World War Two, who recall the hypocrisy and exploitation with which female workers were treated at that time.[115] The filmmakers contrast the interviewees' testimony with propaganda images of "the noble icon of the woman worker as seen in forties newsreels".[116]

Errol Morris' *The Thin Blue Line* uses images to question the veracity of verbal testimony — the inverse of the strategy used in *The Life and Times of Rosie the Riveter*.[117] Morris casts doubt on the words of key witnesses in a murder trial, by juxtaposing their "interviews with images that affirm or undercut what is said, in a spirit of critical irony."[118]

In each of these films, the viewer is confronted with a conflict or dialectic between what they see and what they are told. The viewer must recognise that the two sources of information are in conflict, and judge if one source seems more reliable than the other. This is a more complex relationship than is found in evidentiary editing. We might expect that the top-down influence of language on imagery may

[114] de-Witt et al., "Predictive Coding and the Neural Response to Predictable Stimuli," 8702; Kok et al., "Prior Expectations Bias Sensory Representations in Visual Cortex." 16275.
[115] Field, *The Life and Times of Rosie the Riveter*.
[116] Nichols, "The Voice of Documentary," 21.
[117] Morris, *The Thin Blue Line*.
[118] Nichols, *Introduction to Documentary*, 75–76.

be diminished in these films, as the images do not directly illustrate the spoken words[119].

A third relationship between the two content streams can be observed in films in which there is an ironic or indirect relationship between images and the spoken word, rather than direct contradiction. A contemporary exponent of this style of filmmaking is Adam Curtis. Using his access to the vast archives of the BBC, he has developed a style that combines his voiceover narration with a dense collage of found footage. Describing Curtis' *The Power of Nightmares*, David Bordwell observes:

> The picture track often works on us less as supporting evidence than as a stream of associations, the way metaphors or analogies give thrust to a persuasive speech.[120]

Curtis claims that his use of images was born of "desperation", due to the difficulties of illustrating the economic and political content of his television series in a compelling way.[121] It was only after reaching a dead end using more conventional approaches that he stumbled on the editing style he has become known for. Chris Darke characterises Curtis' "trademark style":

> Mad crash-edits of archival footage vie with ominous tracking shots through deserted institutional corridors, and wide-eyed animals (Curtis favors owls, mantises, and marmots) gaze at heads of states rationally planning the next policy calamity, while snatches of pop music and electronica add a further sonic counterpoint to the image avalanche.[122]

Curtis regards himself as a journalist rather than a documentary filmmaker, and insists on the primacy of spoken narration. He argues that "provided your writing is strong — the words are terribly important", you can "just throw in anything you like" in the way of images. He even expresses amused contempt for film crews since

[119] It is interesting to note that both of the above documentaries involve a critical revision of past events. In these films, the conflict between interpretations of history are expressed as a conflict between AV imagery and the spoken word.
[120] Bordwell and Thompson, "Observations on Film Art: Showing What Can't Be Filmed."
[121] Eaves and Marlow, "Adam Curtis: 'I'm a Modern Journalist.'"
[122] Darke, "Systems Analyst," 23–24.

"they believe that pictures are more important than words and they always want to go to restaurants and get fed".[123]

Of course Curtis is being humorous and provocative; he may get on well with film crews, and he does not choose his images merely at random. The relation of imagery to narration in his films is indirect, his choice of images is aimed primarily at building texture and emotional resonance.

Discussing his work in an interview with Hans Ulrich Obrist, Curtis says:

> Mood-wise I do try and take factual stuff and make it feel like a novel. It doesn't mean it's fiction… but I always want it to have the feel you get from reading a novel, that draws you in emotionally.[124]

Bordwell remarks that "Curtis' images are enigmatic, tangential, or metaphorical", and that the relationship between the edited images also contributes to the mood of his programs. "The ominous, not to say paranoid, tenor of the sequence is aided by the unexpected juxtapositions."[125]

Curtis' comments do seem to suggest that he believes images have negligible epistemic value. Voiceover is the dominant organising force in his programs, and he chooses his images to modify the emotional resonance of his verbal arguments, as a form of visual rhetoric. As entertaining and effective as his documentaries are, in this sense they are a continuation of, even an intensification of the expository mode.

The Influence of Image Duration On Bottom-Up Perception

Curtis frequently uses shots of very brief duration, edited together in often unpredictable ways, using a montage style influenced by music videos. It seems likely that images shown for very short periods reduce the extent of bottom-up processing in comparison to images of longer duration. This is because the influence of automatic, identificatory top-down processes is likely to be greatest in the initial moments of perceptual processing, as the perceiver orientates themselves towards a new event or stimulus. As noted earlier, top-down perceptual processing confers an

[123] Eaves and Marlow, "Adam Curtis: 'I'm a Modern Journalist.'"
[124] Obrist, "In Conversation with Adam Curtis, Part I."
[125] Bordwell and Thompson, "Observations on Film Art : Showing What Can't Be Filmed."

adaptive advantage through improved speed of recognition and response, by pre-empting and curtailing slower, more detailed bottom-up perceptual processes. In order to realise that speed advantage, top-down processes must operate at the earliest stages of perception. This is confirmed by the fMRI studies referred to earlier.[126] We might therefore expect that the influence of those top-down processes will be at its highest in the first instant an image appears on screen, and wane as image duration extends.

As a result, a quick-cutting editing style will tend to maximise the influence of top-down processes on viewers' perception of individual shots. This was one factor in my adoption of a relatively long-take form, which I discuss in the next chapter.

Chapter Summary

The observations and analysis by film theorists such as Michel Chion and Bill Nichols clearly accords with the evidence from the cognitive sciences. In non-fiction films, just as in our external environments, language tends to strongly influence perception, "guiding and structuring our vision".[127] This effect is strongest in the most widespread non-fiction film structure, the expository mode, in which AV imagery serves to illustrate or provide evidence for the voiceover narration. However, there are strategies that may reduce the influence of language over AV imagery, thereby reducing the linguistic attenuation of bottom-up cognition. The next chapter discusses my approach to developing such a strategy.

[126] de-Witt et al., "Predictive Coding and the Neural Response to Predictable Stimuli," 8702; Kok et al., "Prior Expectations Bias Sensory Representations in Visual Cortex." 16275.
[127] Chion, *Audio-Vision*, 7.

5. The Video Project

The evidence I have reviewed in previous chapters indicates that a) language and perception are structurally different ways of knowing, and b) language, though epistemically derivative of embodied experience, exerts a significant influence over perceptual experience. This evidence led me to focus my research on an exploration of the relationship between voiceover narration and AV imagery in non-fiction film.

My approach was to develop a form in which the voiceover and AV imagery tracks were largely independent, each addressing the concerns of the project, but having only an indirect relationship with the other. This form may be expected to reduce the effects of language on perception. In my opinion, this form can produce a viewing experience which is quite distinct from the expository mode. The viewer can observe the AV imagery in a manner we can justifiably expect to have closer similarities to real-world observation. While bottom-up cognition tends to be attenuated by voiceover narration in the expository mode, the film form I am exploring allows bottom-up processes to operate much as they would during real-world observation.

Further, it is probable that a voiceover track which is not directly related to the simultaneous AV imagery track will function as a form of *verbal interference*. This is a technique used by psychologists to interfere with subjects' top-down cognition during experiments. It commonly involves playback of distracting verbal recordings while subjects attempt categorisation and memory tasks, and has been found to reliably interfere with subjects' ability to perform such top-down cognitive tasks.[128] So the disjunction between the verbal and AV imagery tracks in my video project may be expected to interfere with a viewer's self-generated top-down responses to the AV imagery, resulting in the viewer perceiving the imagery with a more "innocent eye" than would be the case if they were to view that AV imagery in the absence of voiceover.

In the context of this program's subject matter, the independence of AV imagery also serves to emphasise the central role of perception in conscious experience. The philosophers Peter Carruthers and Jesse Prinz argue that all consciousness is

[128] Lupyan and Swingley, "Self-Directed Speech Affects Visual Search Performance," 4.

perceptual, and we have no introspective access to our thought processes.[129] According to their Restrictivist philosophies, we become conscious of the outputs of our thought processes only when they are rendered as quasi-perceptual contents (internal speech acts, mental imagery, physical action and so on).

The suggestion that we have little or no introspective access to the workings of our minds developed into a central theme of the video project, with Restrictivism briefly discussed in the voiceover narration. In this context, the independence of AV imagery from voiceover is a strategy to engage the viewer's subjective awareness of their perceiving. The viewer is confronted by AV imagery acting as an independent perceptual and epistemic source.

Overlaps and Correspondences

My creative works consist of two versions of a twenty-eight minute video project. In one version the voiceover narration, spoken by Leonie van Eyck, has been processed in collaboration with the composer Glenn Norman. That post-production work is discussed later in this chapter. The other version of the video project retains the unprocessed voiceover. The AV imagery in each version is identical. The video works can be viewed online, at the locations given in the Creative Works section of this thesis.

The AV imagery consists of recordings of the natural world, shot in Victoria (the western suburbs of Melbourne and the Otway Ranges), and the Northern Territory (around Katherine and Ngukkur). While shooting I chose to avoid the spectacular, instead seeking compositions conducive to observation. My thinking was that such compositions would provide opportunities for the viewer to attend to their processes of perception.

Audio contributes a large part of the perceptual information in AV recordings, just as it contributes significantly to our embodied understanding of our environment. For example, several of the AV recordings convey the presence of insects only through sound, there are none visible. My aim during post-production was to maintain perceptual realism of the AV imagery, both audio and video. The diegetic sound has been cleaned of artefacts introduced during recording (such as wind noise), but

[129] Carruthers, *The Opacity of Mind : An Integrative Theory of Self-Knowledge*; Prinz, *The Conscious Brain*.

there has been little post-production beyond that. Glenn and I attempted to keep the relative loudness of diegetic sound at subjectively realistic levels, while maintaining the intelligibility of the voiceover narration.

Following the reasoning discussed in Chapter Four, I approached the voice over and AV imagery as separate information streams. For the most part, the narration discusses cognitive science, including psychological experiments and philosophical theories. The AV imagery consists of uninflected, relatively long static shots that invite the viewer to observe the recorded landscape before them, and act as a constant intimation of the intrinsic role perception plays in consciousness.

Thus the two content streams investigate the questions surrounding conscious experience in different registers. The one is largely a discussion of and response to external events and theories about consciousness, the other is an invitation to the viewer to reflect on their own perceptual consciousness.

While the two streams address the subject matter of the project using different approaches, there are overlaps in content. In the context of the video project, the recorded landscapes may be suggestive of the mind as an outcome of biological structures and processes. The movement we see and hear — of animals, water and wind — plays a role in emphasising perception, as I will discuss later in this chapter. At the same time, movement draws attention to the innumerable agents and forces that constitute the ecological systems of these landscapes. By analogy, viewers may be reminded of the myriad processes and agents that constitute the human mind.

Some sections of the voiceover narration also discuss parallels between the structures and processes observable in landscapes and those of the mind. In describing multi-agent systems the narration makes reference to the bottom-up, decentralised organisation of social insects (present in the AV sequences as ants and termite mounds), and to the dynamic stability of ecologies.

Thus there is an overlap in the content of the two streams of information. It could be argued that in order to achieve a strict separation between the two streams, this overlap should have been avoided, by removing such references from the voiceover narration.

In making decisions regarding the relationship between particular sequences of AV imagery and concurrent voiceover, I was guided by a sense of emotional resonance between the two streams. I did not intentionally seek metaphors or correspondences. To my mind, doing so may have undermined the independence of the two content streams. Nevertheless, correspondences may occasionally be inferred. As an example, when the narration discusses insight problem solving, the concurrent AV imagery is of gnats in a garden. Viewers may interpret the unpredictable flight paths of these insects as a visual metaphor for the elusive thought processes described in the narration. While I did not deliberately seek such correspondences, once aware of them I did not necessarily reject them, except where I found them insistent and distracting. Again, it might be argued that in a rigorous experiment on the separation of AV imagery and voiceover narration, such correspondences should have been avoided.

I do not believe that the content overlaps and correspondences discussed above are sufficiently direct either to amount to an evidentiary relationship, or to result in linguistic attenuation of viewers' bottom-up cognition. The AV imagery does not illustrate, or provide evidence for the narration. Viewers are confronted by AV imagery which is not simultaneously contextualised and interpreted by the voiceover narration. While viewers may respond by applying their own top-down interpretations of the AV imagery, I believe the greater part of the AV imagery offers little support for such interpretation. Instead, there is an indirect echo resonating between the two streams. I feel this echo enhances the suggestive, associative potential of the otherwise separate content streams.

My video works represent just one approach to an interrogation of the relationship between audio-visual (AV) imagery and voiceover narration in non-fiction films. There is much regarding this relationship that remains to be investigated, and a great deal of potential for filmmakers to explore the wider relationship between perception and language in human cognition and consciousness.

The Viewer as Homunculus

I started this project with a background in educational DVD production, and an intention to make a film about various aspects of contemporary philosophy of mind and cognitive science that I believe have wide significance. An early concern of mine

was to identify possible strategies for effectively communicating content of this level of abstraction in a non-fiction film. It was clear that language, either as narration or interview, would be necessary to convey much of the cognitive science under discussion.

Given that I expected language would carry much of the burden of conveying this content, I began to question the role of AV imagery in the project. I considered the various possibilities of conventional illustration, then of analogy and metaphor. As I thought through potential video sequences, it gradually became apparent that AV imagery can very easily be interpreted as a representation of the contents of (someone's) perception, rather than a representation of an external scene. That is, AV imagery can be taken to be representing what Antonio Damasio has referred to as the "movie-in-the-brain", rather than representing the world.[130]

Many fiction films deliberately invoke such an interpretation, so that the screen becomes a window into a character's subjective experience. To achieve this, filmmakers use a number of impressionistic techniques such as hand-held camera, point-of-view shots, distorting lenses, filtered audio, stream-of-consciousness associative editing structures and the like.

In fiction films these are entirely legitimate techniques, but I felt that in the context of this project's subject matter, such a subjectivist interpretation would be problematic. Representing the screen as a window into subjective experience seemed to me to inevitably evoke "the perennially attractive, but incoherent, model of conscious experience" Daniel Dennett has dubbed Cartesian Materialism. This is "the idea that … 'everything comes together' in some privileged central place in the brain … for 'presentation' to the inner self or homunculus."[131]

Dennett has singled out Cartesian Materialism as "the most tenacious bad idea bedeviling our attempts to think about consciousness".[132] Its tenacity is due to its apparent naturalness, the ease with which we are disposed to think of our self as the observer of the movie-in-the-brain, the already-conscious homunculus to whom the contents of consciousness are presented. Perhaps this apparent naturalness is the legacy of the centuries-long influence of dualistic theories of mind.

[130] Damasio, *Looking for Spinoza : Joy, Sorrow, and the Feeling Brain*, 198.
[131] Dennett and Akins, "Multiple Drafts Model."
[132] Dennett, *Consciousness Explained*, 108.

AV imagery can easily be interpreted from this Cartesian Materialist perspective — the camera becomes the mind's eye, the viewer the homunculus. Given the content of my program, I felt any suggestion of Cartesian Materialism would be misleading and potentially confuse the arguments being discussed.

It may be that any communicative medium, to the extent that it supports subjective representations, can be taken to be a representation of consciousness. Nonetheless, I decided to avoid the use of filmmaking strategies which overtly position the viewer as homunculus, such as the impressionistic techniques mentioned above.

The Long Take

In "When Less Is Less", David MacDougall notes:

> Most of the shots in contemporary documentary films and television programs are only a few seconds long. This is in marked contrast to fiction films and television dramas, in which whole scenes are played out in a single shot. Documentary thus finds itself in the curious company of television commercials and music videos in seeking to maintain audience interest through the dynamics and variety of quick cutting.[133]

MacDougall suggests this is because documentary film producers "are terrified of audience impatience".[134] He argues that the process of tightening a program over several edits "progressively centers particular meanings", thereby increasing coherence and "the economy of signification of the film", but at the cost of a loss of "interpretive space" and context.[135]

I suspect the reduced epistemic role of imagery in evidentiary editing also motivates the use of short shots. Editors become accustomed to using footage less as a vehicle of information than as material to fill gaps in a sequence (the duration of those gaps is determined of course by the length of the voiceover). Under these circumstances, footage may be used primarily as a form of visual stimulus — and the "dynamics and

[133] MacDougall, "When Less Is Less," 36.
[134] Ibid., 38.
[135] Ibid., 41.

variety" of rapid editing provide just such a stimulus. This recalls Serge Daney's description of the image left in a state "of frantic disinheritance."[136]

Voiceover narration and short-duration shots may each play a role in restricting bottom-up cognitive processes. As discussed earlier, the influence of automatic top-down processes is likely to be greatest in the initial moments of perceptual processing, as the perceiver orientates themselves towards the new stimulus. This has clear parallels with MacDougall's observation:

> Like a spark or a stab of lightning, a shot discharges most of its meaning at once, within the first few microseconds of appearing on the screen. If we close our eyes after that first instant, the meaning survives. The mind arrests it like the shutter of a camera. What follows in our response may be very different—a sudden adherence to something happening within the shot, or a kind of coasting perusal. Or so it can be if the shot continues.[137]

MacDougall's observation accords with the hypothesis that over the course of a longer shot, the influence of top-down processes wanes. MacDougall refers to this later perceptual state as "digressive search", a phase when we "inspect details which escape the film's inscriptions of meaning".[138] During this time, our perceptual processes continue to acquire information from the image, though the early top-down processes of recognition and categorisation have completed. It is at this later stage that we can absorb further details, which become the grist to the mill of our bottom-up cognitive processes. This recalls the discussion in Chapter Two regarding the painter John Constable.

Experiments in Watching

I initially chose to record relatively long shots of the landscapes I was shooting, in order to capture unpredictable small events such as the movement of water, of leaves and grasses in the wind, the entrance and exit of flying insects, changes of light. Most recordings were two minutes or longer.

[136] Daney, "Back to Voice," 19.
[137] MacDougall, "When Less Is Less," 37.
[138] Ibid., 43.

My subsequent research uncovered theorists who have discussed observational approaches to recording AV imagery in fiction films. In *Metaphysics of the 'Long Take'*, LeFanu distinguishes between two approaches to the long take. He describes the baroque virtuosity of set-pieces in fiction films (with bravura camera movement used as a marker of directorial style), and opposes that approach to the "contemplative engagement" that marks the long takes of Mizoguchi, Tarkovsky, Hou Hsiao-Hsien and others. In this approach, "the camera is not so much the star, but a kind of self-effacing servant, biding its time", waiting for "the unplanned moment".[139]

To my mind, the unplanned outcomes of these "experiments in *watching*", as LeFanu describes the work of these filmmakers, relate to a form of real-world observation in which we attend without a specific search object or goal. Though "undirected", such observation can lead to an awareness of previously unnoticed features, or a new understanding of the underlying patterns of relationships and structures in the observed. Observation provides the source material for our bottom-up cognitive processes which, beneath consciousness, constantly seek patterns and correspondences. The sudden, unanticipated crystallisation of awareness that can result recalls the insight problem-solving studied by Jonathan Schooler and others.

Chion and the "Microstructure of the Present"

In *Audio-Vision: Sound on Screen*, Chion discusses the use of *visual microrhythms*, such as curls of smoke, rain, or falling petals, seen in the films of Kurosawa, Syberberg, de Oliveira and Tarkovsky. These instil "a vibrating, trembling temporality in the image itself".[140] These small, often unremarkable events "reconstitute the texture of the present".[141]

Such *microevents* are almost always present in recordings of the natural world — small, sporadic events that engender a sense of unfolding time. In the context of this project, this temporal unfolding both emphasises the process of perception, and as representations of natural dynamics, reflect a model of the mind as composed of innumerable small processes and events.

[139] LeFanu, "Metaphysics of the 'Long Take': Some Post-Bazinian Reflections."
[140] Chion, *Audio-Vision: Sound on Screen*, 16.
[141] Ibid., 19.

This is not an argument that only passive observation can stimulate bottom-up cognitive processes — though it's possible the static shots in my video works may seem to suggest that. In real-world perception, we move through our environment and physically interact with it. My choice of static shots was a practical one. The production constraints of carrying all recording equipment, sometimes for considerable distances, meant that I took only a tripod or monopod for stabilisation. Smooth travelling shots were not possible. As discussed previously, though hand-held or head-mounted camera shots may signify documentary realism, in my view they are not perceptually realistic, since the camera movements are not offset by the viewer's proprioceptive feedback mechanisms. The static long take was the most perceptually realistic approach to image acquisition available.

When I came to edit the AV imagery, several factors influenced my approach, which increasingly gravitated towards relatively long shots. These included the observations above regarding:

a). exhausting the initial wave of top-down processing which dominates short-duration shots,

b). foregrounding perception as a core component of consciousness, and as a distinct way of knowing and thinking (distinct from linguistic knowledge and cognition). The unfolding of time perceptible in “microevents” over the course of relatively long shots emphasises the act of perception.

c). reducing the increased cognitive load which results from separating the tracks of voiceover narration and imagery. Viewers of early edits reported that the effort to process the voiceover and imagery as largely independent sources of information resulted in a feeling of increased cognitive load. They reported slipping in and out of focus from one of the tracks to the other. Longer takes, in combination with a slower pace of narration, may reduce the feeling that information is slipping past.

This feeling of increased cognitive load suggests the extent to which conventional evidentiary editing may reduce that load. According to the evidence presented here, voiceover narration in the expository mode feeds into the viewer's top-down

cognitive processes and attenuates their bottom-up processing of AV imagery. That attenuation of mental processing may result in a feeling of reduced cognitive load.[142]

The Long Take in Non-Fiction Film

While contemporary documentary filmmaking may be dominated by a quick-cutting editing style, as discussed by MacDougall, several non-fiction filmmakers have used relatively long takes. Two contemporary exponents are James Benning and Tacita Dean. Both have made several films that are oriented towards the observation of place over time. Benning's *13 Lakes* consists of thirteen ten-minute static shots of lakes in the United States.[143] Dean's *Fernsehturm* was shot in a revolving restaurant in the Berlin television tower, capturing the changing light as time passes and day fades to night.[144] Jeanette Winterson wrote that *Fernsehturm* engenders "a different experience of time, and a different relationship to everyday objects."[145] Scott MacDonald discusses "the viewer's growing awareness of his or her own perceptiveness" as they "measure the subtle changes that occur within each shot over time" in Benning's *13 Lakes.*[146] Benning describes his method as "looking and listening" and claims that "place can only be understood over time; that is, that place is a function of time."[147]

My film was shot in several locations; it is not about a particular place. But there are parallels with the films cited above, in the focus on perception and observation, and the use of shot duration as a method of involving the viewer in that focus, drawing attention to the viewer's own processes of perception.

The Acousmatic Voice

To this point I have discussed voiceover narration only in the context of its role as a top-down linguistic influence on perception and bottom-up cognition. Another aspect of the voiceover that influenced the development of my video works is its

[142] I would note, however, that psychologists describe bottom-up cognition as being largely effortless. Since we are not conscious of bottom-up cognition, we cannot ascribe feelings of effort to it. This does not preclude the probability that non-conscious thought requires mental resources. That resource usage may contribute to a conscious feeling of cognitive load when simultaneously trying to follow a verbal narration.
[143] Benning, *13 Lakes*.
[144] Dean, *Fernsehturm*.
[145] Winterson, "Much Ado About Nothing."
[146] MacDonald, "Testing Your Patience."
[147] Ibid.

relation to the subjectively experienced inner voice. Relevant here is Michel Chion's discussion of the *acousmatic voice*, the recorded voice which is heard without its source (the speaker's body) being seen. A sub-category of the acousmatic voice is the *I-voice*. Chion claims viewers involuntarily identify with an I-voice as if it were their own.[148] The defining quality of the I-voice is its sense of intimacy, which is achieved by close miking, lack of reverberation, and non-declamatory style. The result of these technical and performative qualities, according to Chion, is that the I-voice compels the identification of the viewer, since its perspective is that of the viewer's own voice.

This is an interesting claim when viewed in the light of the Restrictivist argument that we are conscious of the contents of our thought processes only when they are internally rendered in quasi-perceptual form, mostly commonly as inner speech. On this view, our inner speech consists of the outcomes of otherwise non-conscious thought processes, and we have no introspective access to its multiple sub-personal sources. Considering that lack of access in the light of Chion's argument, it would seem that voiceover narration delivered in the style of the I-voice may have a subjective similarity to inner speech. Though the timbre and manner of speech of a particular voiceover may not match our own, though the content of the narration may not be as familiar as the content of our own inner speech, perhaps the intimate perspective of certain styles of narration may encourage us to accept their contents almost as if they were indeed the contents of our own thoughts, rendered in perceptual form?

Chion's discussion of the I-voice is couched in a kind of poetic dualism. He bases much of his argument in *The Voice In Cinema* on the suggestion that "the voice enjoys a certain proximity to the soul, the shadow, the double — these immaterial, detachable representations of the body, which survive its death and sometimes even leave it during life".[149] The core metaphor he uses throughout this book to suggest the power of the acousmêtre is spiritual possession, the bodiless voice as ghost.

However, it is possible that Chion's observations may find more plausible materialist explanations. For example, there is speculation from some psychologists that the auditory verbal hallucinations that are symptomatic of schizophrenia may

[148] Chion, *The Voice in Cinema*, 50–51.
[149] Ibid., 47.

be the result of an inability to recognise some inner speech acts as self-generated (for a discussion, see Wu, 2013).[150] These psychologists argue that in healthy subjects, inner speech is tagged as self-generated, using a self-monitoring mechanism similar to corollary discharge. Failure of this self-monitoring mechanism in schizophrenia can mean subjects lose track of some inner speech acts as self-generated, and take them to be the voice of another. Other researchers, including Wu, argue that auditory verbal hallucinations are the result of over-activity in the auditory cortex, and have the aural qualities of another's voice. While these theories are still a matter of debate among psychologists, they raise questions as to the mechanisms with which we recognise inner speech as our own, and by extension, the permeability of inner speech by acousmatic speech in AV media, and by voiceover narration in particular.

It remains an open possibility that the intimate, bodiless voice of a voiceover narrator can partially evade an attribution of otherness, and diminish the critical distance a viewer may bring to bear on the speech of an observed presenter.

I was interested in pursuing an intimate style of voiceover for two reasons. I wanted the narrator's voice to suggest a personal engagement with the theories under discussion, since I believe their implications are relevant to us at a personal level. As a result, I tried to achieve a degree of intimacy in the texture and delivery of the narration, through choice of narrator and through direction. Secondly, I believed that if a suggestion of the inner voice could be established, it might open up possibilities of exploring the model of mind that the project presents, through formal means. For reasons of brevity and clarity, I chose not to emulate the halting, meandering, episodic and repetitive qualities that seem often to characterise our inner speech. Instead, I decided to experiment with the possibilities that may inhere in the fragmentation of the voiceover.

The Fragmented Voice

A central theme of the video project is our lack of introspective access, and our inability to recognise this lack. Indeed 'lack of insight' inadequately expresses our condition, in which we are not even aware of the dimensions of our lack, and which

[150] Wu, "Self-Monitoring and Auditory Verbal Hallucinations in Schizophrenia."

might better be expressed by the medical term *neglect*, or the Rumsfeldian "unknown unknowns".[151]

This philosophical theme found expression in the fragmentation of the voiceover, in which unintelligible voice fragments underlie the central thread of the narration, as if competing with it, alternative viewpoints which threaten to emerge and supplant it. These fragments may suggest the multiple, semi-autonomous processes which are referred to throughout the video, the origin of many of our thoughts and behaviours, but inaccessible to us. Their sometimes disruptive jostling, as they each attempt to promote themselves and supplant others, are reflected in the often-unintelligible voice fragments which emerge and disappear, recalling the dropouts and echoes of a transmission from a distant and inaccessible source.[152]

At the same time, this fragmentation explores two other possibilities. The unintelligibility of the fragments may reduce the perceived stability and narrative coherence of the voiceover, thereby diminishing its dominance over AV imagery as the film's primary epistemic source. Secondly, it emphasises the physical, non-linguistic aspect of vocalisation. In keeping with the focus on perception, I was interested in foregrounding the physical aspect of the voice, and so retained the sounds of breathing, the swallows and mouth noises that an audio editor would usually remove. The processing of the voice into fragments also emphasises the non-linguistic aspect of the voice, by transforming the voice into sounds without intelligible linguistic content.

Michel Chion writes that the I-voice need not contain words. He cites the breathing of the astronaut at the end of Kubrick's *2001*, and an early scene in the Lynch's *The Elephant Man*, where the face of John Merrick is covered, but we can hear his breathing and painful swallowing.[153]

---

[151] "There are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know." Donald Rumsfeld, then U.S. Secretary of Defense, in reply to a question at a U.S. Department of Defense news briefing on February 12, 2002. Retrieved from http://en.wikipedia.org/wiki/There_are_known_knowns

[152] The fragmentation of the voiceover recalls the early neural network speech synthesiser NETtalk, developed by Terry Sejnowski and Charles Rosenberg in the 1980s. Audio from the neural network is available online at https://youtu.be/gakJlr3GecE from the contributor Frawstbyte. The three recordings of random, incoherent fragments of voice gradually coalescing into intelligible phrases, is a striking example of the learning process of a bottom-up neural network — and by analogy, suggestive of the bottom-up emergence of our conscious thoughts from sub-personal systems.

[153] Kubrick, *2001*; Lynch, *The Elephant Man*.

> We might call this an effect of corporeal implication, or involvement of the spectator's body, when the voice makes us feel in our body the vibration of the body of the other, of the character who serves as a vehicle for the identification. The extreme case of corporeal implication occurs when there is no dialogue or words, but only closely present breathing or groans or sighs. We often have as much difficulty distancing ourselves from this to the degree that the sex, age, and identity of the one who thus breathes, groans, and suffers aren't marked in the voice. It could be me, you, he, she.[154]

I worked with the composer Glenn Norman in experimenting with several different approaches to fragmentation of the voice, gradually developing a style of fragmentation which sounds more analogue than digital (and so more fragment than audio effect), and which retains what we judged to be sufficient intelligibility of the linguistic content so as not to frustrate the viewer.

The results are included as a separate video piece. Due to the length of the film, the first half of that video has substantially developed fragmentation of the voiceover, while the second half is currently in a preliminary stage.

[154] Chion, *The Voice in Cinema*, 53.

6. Conclusion

On the evidence reviewed here, humans make use of two different modes of cognition, which are represented and processed in the brain using different, though interacting principles. Embodied experience feeds into a bottom-up architecture, enabling pattern-recognition, concept acquisition, and insight problem solving. At the same time, our top-down stored-concept architecture attempts to predict the outcomes of those bottom-up processes, and achieves efficiencies by pre-emptively inhibiting those processes that conform to its predictions. Experimental studies have shown that language is tightly integrated into this top-down architecture, with linguistic effects on colour perception, motion perception, visual search and categorisation.

Language is an enormously powerful tool for cognition. It enables propositional reasoning, and the communication of knowledge across time and space. As Daniel Dennett has written, "The expressive, information-encoding properties of … language are practically limitless (in at least some dimensions)", and "each individual human brain, thanks to its communicative links, is the beneficiary of the cognitive labors of the others in a way that gives it unprecedented powers."[155] This thesis does not dispute those claims (given Dennett's caveat in the first of the quotes above). However, I have argued that despite language's indisputable power, it is clear that it is not the medium of much of the human animal's cognitive activity. Symbolic systems such as language rely on reference to external sources of meaning. In the human brain, those meanings are derived from our embodied interaction with our environments, from perception and action in the world. Our foundational concepts are acquired through the ability of our bottom-up cognitive processes to recognise regularities in the information flows from our embodied interaction with our environments.

Though the meanings of language are derivative of perception and action, experimental evidence shows the influence of language on perception. It acquires this influence by being tightly integrated into phylogenetically older top-down processes, such as predictive processing. As a result of its integration, language

[155] Dennett, "The Role of Language in Intelligence".

exhibits the ability shown by other top-down processes, of pre-empting and attenuating bottom-up cognition.

The relationship between AV imagery and voiceover narration in non-fiction films is a particular case of the relationship between perception and language in human cognition. Though the influence of de Saussure's semiology has resulted in some theorists considering AV imagery to be a text, the meanings of AV imagery are not conventional in the way the meanings of language and other symbolic systems are. AV imagery is perceptually realistic, and so is available to bottom-up cognitive processes in much the same way as real-world perception. A corollary of this realism is that linguistic expression, such as voiceover narration, influences our perception of AV imagery, just as it has been shown to influence real-world perception. Film theorists have observed the influence of language on our perception of AV imagery in non-fiction films, particularly where a film conforms to the structure identified by Bill Nichols as the expository mode. To the extent that we are told what we are looking at, the less we may see.

The influence of top-down processes tends to be greatest in the earliest moments of perceptual processing, and so is maximised by the use of short shots. MacDougall, LeFanu and others have noted the changes in viewing experience that emerge over shots of longer duration. I have suggested this changed experience results from the exhaustion of the initial wave of top-down predictive recognition processes.

My creative works experiment with an alternative to the direct relationship between AV imagery and voiceover narration characteristic of the expository mode. Rather than illustrating, or providing evidence for the content of the voiceover, the AV imagery in my videos operates as an independent epistemic source. This structure, in combination with the use of relatively long shots, foregrounds the act of perception. That emphasis on perception reflects both its fundamental role in cognition, and the Restrictivist philosophy of perceptual consciousness. On the Restrictivist view discussed in the videos, we have no introspective access to our thought processes, and become aware of those thoughts only when their outputs are rendered in quasi-perceptual form. The Restrictivist argument situates perceptual experience as the explanandum of consciousness research.

The sound composer Glenn Norman and I experimented with the fragmentation of the voiceover narration in the video works, to emphasise the voice's perceptual qualities, and to suggest our lack of access to our thought processes. The innumerable vocal fragments reflect an image of the mind as composed of multiple sub-personal processes. These semi-autonomous agents must often compete for a temporary claim on the limited resources of consciousness, to achieve the downstream effects of consciousness: influence over behaviour, memory and other thought processes.

In these ways, the form of my creative works expresses the philosophical concerns that also dominate the content of the voiceover narration — a model of the mind consisting of multiple, semi-autonomous agents, and our lack of introspective access to the sources of the thoughts we perceive. At the same time, the video project's structure is a response to the scientific evidence and theoretical work presented in this thesis regarding the relationship between perception and language, and the implications of this for non-fiction films employing voiceover narration. On the evidence presented here, the form I have used can be expected to diminish the influence of voiceover narration over viewers' perception of AV imagery. This may result in an altered experience for viewers, with an expanded epistemic role for AV imagery, emphasising the act of perception as central to questions on the nature of consciousness.

Bibliography

Alogna, V. K., M. K. Attaya, P. Aucoin, Š. Bahnik, S. Birch, A. R. Birt, B. H. Bornstein, et al. "Registered Replication Report: Schooler and Engstler-Schooler (1990)." *Perspectives on Psychological Science* 9, no. 5 (September 2014): 556–78.

Anderson, Michael L. "Embodied Cognition: A Field Guide." *Artificial Intelligence* 149, no. 1 (2003): 91.

Andersson, Richard, Fernanda Ferreira, and John M. Henderson. "I See What You're Saying: The Integration of Complex Speech and Scenes during Language Comprehension." *Acta Psychologica* 137, no. 2 (June 2011): 208–16. doi:10.1016/j.actpsy.2011.01.007.

Aydede, Murat, and Brian McLaughlin. "The Language of Thought Hypothesis." Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Fall 2010. <http://plato.stanford.edu/archives/fall2010/entries/language-thought/>.

Bazin, André. *What Is Cinema?*. Translated by Hugh Gray. Vol. 1. 2 vols. Berkeley: University of California Press, 1967.

Benning, James. *13 Lakes*. 16mm negative, 35mm print, Documentary, 2004.

Boden, Margaret A. *Mind as Machine : A History of Cognitive Science.* Vol. 1. 2 vols. New York: Oxford University Press, 2006.

Bordwell, David, and Kristin Thompson. "Observations on Film Art : Showing What Can't Be Filmed." *David Bordwell's Website on Cinema*, 2009. http://www.davidbordwell.net/blog/2009/03/04/showing-what-cant-be-filmed/.

Bos, Maarten W., and Ap Dijksterhuis. "Unconscious Thought Works Bottom-Up and Conscious Thought Works Top-Down When Forming an Impression." *Social Cognition* 29, no. 6 (2011): 727–37.

Brakhage, Stan. "From *Metaphors on Vision*." In *The Avant-Garde Film: A Reader of Theory and Criticism*, edited by P. Adams Sitney, 120–28. Anthology Film Archives 3. New York: New York University Press, 1978.

Brown, Charity, and Toby J. Lloyd-Jones. “Verbal Overshadowing of Multiple Face and Car Recognition: Effects of within- versus across-Category Verbal Descriptions.” *Applied Cognitive Psychology* 17, no. 2 (March 2003): 183–201. doi:10.1002/acp.861.

Calvert, Gemma, Charles Spence, and Barry E. Stein. *The Handbook of Multisensory Processes. [electronic resource]*. Cambridge, Mass.: MIT Press, 2004. http://cognet.mit.edu.ezp.lib.unimelb.edu.au/erefs/handbook-of-multisensory-processes.

Carruthers, Peter. *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. Oxford; New York: Oxford University Press, 2006.

———. *The Opacity of Mind : An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press, 2011.

Chion, Michel. *Audio-Vision: Sound on Screen*. Edited and translated by Claudia Gorbman. New York: Columbia University Press, 1994.

———. *The Voice in Cinema*. Edited and translated by Claudia Gorbman. New York: Columbia University Press, 1999.

Churchland, Paul M. “Functionalism at Forty: A Critical Retrospective.” *The Journal of Philosophy*, 2005, 33–50.

Clark, Andy. *Associative Engines: Connectionism, Concepts, and Representational Change*. Cambridge, Mass: MIT Press, 1993.

———. Edge Annual Question 2011: What Scientific Concept Would Improve Everybody’s Cognitive Toolkit?, 2011. http://edge.org/response-detail/10404.

———. *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Explorations in Cognitive Science: 6. Cambridge, Mass.: MIT Press, 1989.

———. “Soft Selves and Ecological Control.” In *Distributed Cognition and the Will: Individual Volition and Social Context*, 101–22. Cambridge, MA: MIT Press, 2007.

Currie, Gregory. *Image and Mind: Film, Philosophy and Cognitive Science*. Cambridge: Cambridge University Press, 2008.

Damasio, Antonio R. *Looking for Spinoza : Joy, Sorrow, and the Feeling Brain*. 1st ed. New York: Harcourt, 2003.

Daney, Serge. "Back to Voice: On Voices Over, In, Out, Through." *Cinema Comparat/ive Cinema* 1, no. 3 (2013): 18–20.

Darke, Chris. "Systems Analyst." *Film Comment* 48, no. 4 (August 2012): 22.

Dean, Tacita. *Fernsehturm*. 16mm anamorphic, 2001.

Dennett, Daniel C. *Consciousness Explained*. Boston: Little, Brown and Co., 1991.

———. "The Role of Language in Intelligence." Cognitive Science ePrint Archive. *CogPrints*, 1994. http://cogprints.org/192/1/rolelang.htm.

Dennett, Daniel C., and Kathleen Akins. "Multiple Drafts Model." *Scholarpedia* 3, no. 4 (2008): 4321. doi:10.4249/scholarpedia.4321.

De Saussure, Ferdinand. *Course in General Linguistics.* Translated by Roy Harris. Bloomsbury Revelations. London: Bloomsbury Publishing, 2013.

Descartes, René. *Discourse on Method, Optics, Geometry, and Meteorology.* Translated by Paul J Olscamp. Library of Liberal Arts 211. Indianapolis: Bobbs-Merrill, 1965.

de-Wit, L., B. Machilsen, and T. Putzeys. "Predictive Coding and the Neural Response to Predictable Stimuli." *Journal of Neuroscience* 30, no. 26 (June 30, 2010): 8702–3. doi:10.1523/JNEUROSCI.2248-10.2010.

Dijksterhuis, Ap, Maarten W. Bos, Loran F. Nordgren, and Rick B. van Baaren. "On Making the Right Choice: The Deliberation-Without-Attention Effect." *Science* 311, no. 5763 (February 17, 2006): 1005–7.

Dijksterhuis, Ap, and Loran F. Nordgren. "A Theory of Unconscious Thought." *Perspectives on Psychological Science* 1, no. 2 (2006): 95–109.

Dreyfus, Hubert L. *What Computers Can't Do: A Critique of Artificial Reason.* New York: Harper & Row, 1972.

Eagleman, David. *Incognito: The Secret Lives of Brains*. New York: Pantheon Books, 2011.

Eaves, Hannah, and Jonathan Marlow. “Adam Curtis: ‘I’m a Modern Journalist.’” *Greencine*, May 30, 2005. http://www.greencine.com/central/node/430?page=0%2C1.

Field, Connie. *The Life and Times of Rosie the Riveter*. 35mm, Documentary. Clarity Films, 1980.

Flaherty, Robert J. *Nanook of the North*. Black and white, 35mm, silent, Documentary. Pathé Exchange, 1922. https://archive.org/details/nanookOfTheNorth1922.

Fodor, Jerry A. *The Language of Thought.* Language & Thought Series. New York: Crowell, 1975.

Fogassi, Leonardo, and Vittorio Gallese. “Action as a Binding Key to Multisensory Integration.” In *The Handbook of Multisensory Processes*, edited by Gemma A. Calvert, Charles Spence, and Barry E. Stein, 425–41. Cambridge, Mass.: MIT Press, 2004.

Fukushima, Kunihiko. “Artificial Vision by Multi-Layered Neural Networks: Neocognitron and Its Advances.” *Neural Networks* 37 (January 2013): 103–19. doi:10.1016/j.neunet.2012.09.016.

Gallese, Vittorio, and Michele Guerra. “Embodying Movies: Embodied Simulation and Film Studies.” *Cinema : Journal of Philosophy and the Moving Image* 3 (2012): 183–210.

Gallese, Vittorio, and George Lakoff. “The Brain’s Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge.” *Cognitive Neuropsychology* 22, no. 3–4 (May 2005): 455–79. doi:10.1080/02643290442000310.

Garson, James. “Connectionism.” Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*, 2012. <http://plato.stanford.edu/archives/win2012/entries/connectionism/>.

Gazzaniga, Michael S., and Joseph E. Ledoux. *The Integrated Mind*. New York: Plenum Press, 1978.

Gruber, Howard E. “Insight and Affect in the History of Science.” In *The Nature of Insight.*, edited by Robert J. Sternberg and Janet E. Davidson, 397–431. Cambridge, Mass.: The MIT Press, 1995.

Guerra, Michele. “Film Style: A Motor Approach.” *Society for Cognitive Studies of the Moving Image*, July 13, 2012. http://scsmi-online.org/forum/film-style-a-motor-approach.

Haugeland, John. *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press, 1985.

Heimann, Katrin, Maria Alessandra Umiltà, Michele Guerra, and Vittorio Gallese. “Moving Mirrors: A High-Density EEG Study Investigating the Effect of Camera Movements on Motor Cortex Activation during Action Observation.” *Journal of Cognitive Neuroscience* 26, no. 9 (September 2014): 2087–2101. doi:10.1162/jocn_a_00602.

Hobbes, Thomas. *Leviathan: Or, The Matter, Forme and Power of a Commonwealth Ecclesiasticall and Civil.* Edited by Ian Shapiro. Rethinking the Western Tradition. New Haven, Connecticut; London: Yale University Press, 2010.

Horst, Steven. “Symbols and Computation: A Critique of the Computational Theory of Mind,” 1996. http://shorst.web.wesleyan.edu/papers/ctmarticle.htm.

———. “The Computational Theory of Mind.” Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*, 2011. <http://plato.stanford.edu/archives/spr2011/entries/computational-mind/>.

Jackson, Frank. “Epiphenomenal Qualia.” *The Philosophical Quarterly*, no. 127 (1982): 127.

Kahneman, Daniel, and Amos Tversky. “Choices, Values, and Frames.” *American Psychologist* 39, no. 4 (April 1984): 341–50. doi:10.1037/0003-066X.39.4.341.

Kahneman, Daniel. *Thinking, Fast and Slow*. London: Penguin, 2012.

Kayser, Christoph, Nikos K. Logothetis, and Stefano Panzeri. ‘Visual Enhancement of the Information Representation in Auditory Cortex.’ *Current Biology* 20, no. 1 (January 2010): 19–24. doi:10.1016/j.cub.2009.10.068.

Keil, Frank C., and Robert A. Wilson, eds. *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Mass.: MIT Press, 1999.

Kok, P., D. Rahnev, J. F. M. Jehee, H. C. Lau, and F. P. de Lange. "Attention Reverses the Effect of Prediction in Silencing Sensory Signals." *Cerebral Cortex* 22, no. 9 (November 2, 2011): 2197–2206. doi:10.1093/cercor/bhr310.

Kok, P., G. J. Brouwer, M. A. J. van Gerven, and F. P. de Lange. "Prior Expectations Bias Sensory Representations in Visual Cortex." *Journal of Neuroscience* 33, no. 41 (9 October 2013): 16275–84. doi:10.1523/JNEUROSCI.0742-13.2013.

Kreiman, Gabriel. "Biological Object Recognition." *Scholarpedia* 3, no. 6 (2008): 2667.

Kubrick, Stanley. *2001: A Space Odyssey*. 65mm, Science Fiction. Metro-Goldwyn-Mayer (MGM), 1968.

Lakoff, George, and Mark Johnson. *Metaphors We Live By*. Chicago: University of Chicago Press, 1980.

Lakoff, George, and Mark Johnson. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought.* New York: Basic Books, 1999.

LeFanu, Mark. "Metaphysics of the 'Long Take': Some Post-Bazinian Reflections." *P.O.V.*, no. 4 (December 1997): 7–21.

Levitin, Daniel J. *This Is Your Brain on Music: The Science of a Human Obsession*. New York: Dutton, 2006.

Lewicki, Pawel, Thomas Hill, and Elizabeth Bizot. "Acquisition of Procedural Knowledge about a Pattern of Stimuli That Cannot Be Articulated." *Cognitive Psychology* 20, no. 1 (January 1988): 24–37. doi:10.1016/0010-0285(88)90023-0.

Lupyan, Gary. "From Chair to 'Chair': A Representational Shift Account of Object Labeling Effects on Memory." *Journal of Experimental Psychology: General* 137, no. 2 (2008): 348–69. doi:10.1037/0096-3445.137.2.348.

Lupyan, Gary. "Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis." *Frontiers in Psychology* 3 (2012). doi:10.3389/fpsyg.2012.00054.

Lupyan, Gary, David H. Rakison, and James L. McClelland. “Language Is Not Just for Talking: Redundant Labels Facilitate Learning of Novel Categories.” *Psychological Science* 18, no. 12 (2007): 1077–83.

Lupyan, Gary, and Michael J. Spivey. “Making the Invisible Visible: Verbal but Not Visual Cues Enhance Visual Detection.” *PloS One* 5, no. 7 (2010): e11452.

Lupyan, Gary, and Daniel Swingley. “Self-Directed Speech Affects Visual Search Performance.” *The Quarterly Journal of Experimental Psychology* 65, no. 6 (June 2012): 1068–85. doi:10.1080/17470218.2011.647039.

Lupyan, G., and E. J. Ward. “Language Can Boost Otherwise Unseen Objects into Visual Awareness.” *Proceedings of the National Academy of Sciences* 110, no. 35 (August 12, 2013): 14196–201. doi:10.1073/pnas.1303312110.

Lynch, David. *The Elephant Man*. 35mm, Drama. Paramount Pictures, 1980.

MacDonald, Scott. “Testing Your Patience.” *Artforum International.* 46, no. 1 (September 2007): 429–30,432,434–35,437,494. http://search.proquest.com.ezp.lib.unimelb.edu.au/docview/214351996?accountid=12372

MacDougall, David. “When Less Is Less: The Long Take in Documentary.” *Film Quarterly* 46, no. 2 (December 1992): 36–46. doi:10.2307/1213006.

McCulloch, Warren, and Walter Pitts. “A Logical Calculus of the Ideas Immanent in Nervous Activity.” *Bulletin of Mathematical Biophysics* 5, no. 4 (December 1943): 115.

Metz, Christian. *Film Language: A Semiotics of the Cinema.* Translated by Michael Taylor. New York: Oxford University Press, 1974.

Miller, George A. “The Magical Number Seven, plus or Minus Two: Some Limits on Our Capacity for Processing Information.” *Psychological Review* 63, no. 2 (March 1956): 81–97.

Minsky, Marvin Lee. *The Society of Mind*. New York: Simon and Schuster, 1986.

Morris, Errol. *The Thin Blue Line*. 35mm & Super 16, Documentary. Miramax, 1988.

Nagel, Thomas. “What Is It Like to Be a Bat?” *The Philosophical Review* 83 (1974): 435–50.

NETtalk Test, 2012. https://youtu.be/gakJlr3GecE.

Nichols, Bill. *Introduction to Documentary*. 2nd ed. Bloomington: Indiana University Press, 2010.

———. “The Voice of Documentary.” *Film Quarterly* 36, no. 3 (April 1983): 17–30. doi:10.2307/3697347.

Obrist, Hans Ulrich. “In Conversation with Adam Curtis, Part I.” *E-Flux*, no. 32 (February 2012). http://www.e-flux.com/journal/in-conversation-with-adam-curtis-part-i/.

Peckhaus, Volker. “Leibniz’s Influence on 19th Century Logic.” Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*, 2014. <http://plato.stanford.edu/archives/spr2014/entries/leibniz-logic-influence/>.

Poincaré, Henri. “Mathematical Creation.” *The Monist* no. 3 (1910): 321-335.

Prince, Stephen. “The Discourse of Pictures: Iconicity and Film Studies.” *Film Quarterly*, no. 1 (1993): 16.

Prinz, Jesse. *The Conscious Brain*. Oxford: Oxford University Press, 2012.

Ravassard, Pascal, Ashley Kees, Bernard Willers, David Ho, Daniel A. Aharoni, Jesse Cushman, Zahra M. Aghajan, and Mayank R. Mehta. “Multisensory Control of Hippocampal Spatiotemporal Selectivity.” *Science*, May 2, 2013. doi:10.1126/science.1232655.

Ruskin, John. *The Elements of Drawing: In Three Letters to Beginners.* Project Gutenberg: 30325. Project Gutenberg, 2009. Accessed January 6, 2015. https://www.gutenberg.org/files/30325/30325-h/30325-h.htm.

Scaruffi, Piero. “Artificial Neural Networks.” Online Book. *The Nature of Consciousness: Consciousness, Life and Meaning*. http://www.scaruffi.com/nature/neu01.html.

Schmidhuber, Jürgen. “Deep Learning in Neural Networks: An Overview.” *Neural Networks* 61 (January 2015): 85–117. doi:10.1016/j.neunet.2014.09.003.

Schooler, Jonathan W. “Verbalization Produces a Transfer Inappropriate Processing Shift.” *Applied Cognitive Psychology* 16, no. 8 (December 2002): 989–97. doi:10.1002/acp.930.

Schooler, Jonathan W., and Tonya Y. Engstler-Schooler. “Verbal Overshadowing of Visual Memories: Some Things Are Better Left Unsaid.” *Cognitive Psychology* 22, no. 1 (1990): 36–71.

Schooler, Jonathan W., Stellan Ohlsson, and Kevin Brooks. “Thoughts beyond Words: When Language Overshadows Insight.” *Journal of Experimental Psychology: General* 122, no. 2 (1993): 166.

Shanahan, Murray. “The Frame Problem.” Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, November 22, 2009. <http://plato.stanford.edu/archives/win2009/entries/frame-problem/>.

Sio, Ut Na, and Thomas C. Ormerod. “Does Incubation Enhance Problem Solving? A Meta-Analytic Review.” *Psychological Bulletin* 135, no. 1 (2009): 94–120. doi:10.1037/a0014212.

Sobchack, Vivian. “What My Fingers Knew: The Cinesthetic Subject, or Vision in the Flesh.” In *Carnal Thoughts*, 53–84. Berkeley: University of California Press, 2004.

Taberham, Paul. “Bottom-Up Processing, Entoptic Vision and the Innocent Eye in Stan Brakhage’s Work.” *Projections* 8, no. 1 (June 1, 2014): 1–22. doi:10.3167/proj.2014.080102.

Tversky, Amos, and Daniel Kahneman. “Judgment under Uncertainty: Heuristics and Biases.” *Science*, no. 4157 (1974): 1124–31.

Veta, Mitko, Paul J. van Diest, Stefan M. Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, et al. “Assessment of Algorithms for Mitosis Detection in Breast Cancer Histopathology Images.” *Medical Image Analysis* 20, no. 1 (2015): 237–48. http://www.sciencedirect.com/science/article/pii/S1361841514001807

Wallace, Mark T., Ramnarayan Ramachandran, and Barry E. Stein. "A Revised View of Sensory Cortical Parcellation." *Proceedings of the National Academy of Sciences of the United States of America* 101, no. 7 (2004): 2167–72.

Wallas, Graham. *The Art of Thought*. London: J. Cape, 1926.

Watt, Harry, and Basil Wright. *Night Mail*. 35mm, Documentary, 1936. https://youtu.be/FkLoDg7e_ns.

Wees, William C. *Light Moving in Time: Studies in the Visual Aesthetics of Avant-Garde Film.* Berkeley: University of California Press, 1992.

Wickham, Lee H. V., and Hayley Swift. "Articulatory Suppression Attenuates the Verbal Overshadowing Effect: A Role for Verbal Encoding in Face Identification." *Applied Cognitive Psychology* 20, no. 2 (March 2006): 157–69. doi:10.1002/acp.1176.

Wilson, Timothy D. *Strangers to Ourselves: Discovering the Adaptive Unconscious.* Cambridge, Mass.: Belknap Press of Harvard University Press, 2002.

Wilson, Timothy D., and Jonathan W. Schooler. "Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions." *Journal of Personality and Social Psychology* 60, no. 2 (1991): 181.

Winterson, Jeanette. "Much Ado About Nothing." *The Guardian*, September 29, 2005. http://www.guardian.co.uk/film/2005/sep/29/1.

Wolchover, Natalie. "As Machines Get Smarter, Evidence They Learn Like Us." *Quanta Magazine*, July 23, 2013. https://www.simonsfoundation.org/quanta/20130723-as-machines-get-smarter-evidence-they-learn-like-us/.

Wu, Wayne. "Self-Monitoring and Auditory Verbal Hallucinations in Schizophrenia." *The Brains Blog*, August 16, 2013. http://philosophyofbrains.com/2013/08/16/self-monitoring-and-auditory-verbal-hallucinations-in-schizophrenia.aspx.

Zhou, Yong-Di, and Joaquín M. Fuster. "Visuo-Tactile Cross-Modal Associations in Cortical Somatosensory Cells." *Proceedings of the National Academy of Sciences of the United States of America* 97, no. 17 (2000): 9777–82.

Creative Works

My creative works consist of two video pieces.

The videos are high-definition (resolution of 1920 by 1080 pixels), colour.

Each is of 28 minutes duration.

They can be viewed at the URLs given below:

How It Seems (original voiceover version)

Full URL: `https://vimeo.com/136061110/f84f77447c`

Shortened URL: `bit.ly/1LFN11E`

How It Seems (processed voiceover version)

Full URL: `https://vimeo.com/136062362/7e8dae1449`

Shortened URL: `bit.ly/1WHgjNO`

Appendix: Connectionism and Artificial Neural Networks

Connectionism is an alternative to the symbolic architecture implemented in most computers today. Its roots can be traced back to McCulloch and Pitts' computational model of the neuron in 1943.[156]

Connectionism sets out to emulate aspects of human cognition by implementing a simplified neural structure in computer systems. Thus Connectionist systems are often known as artificial neural networks.

A Brief Description of an Artificial Neural Network

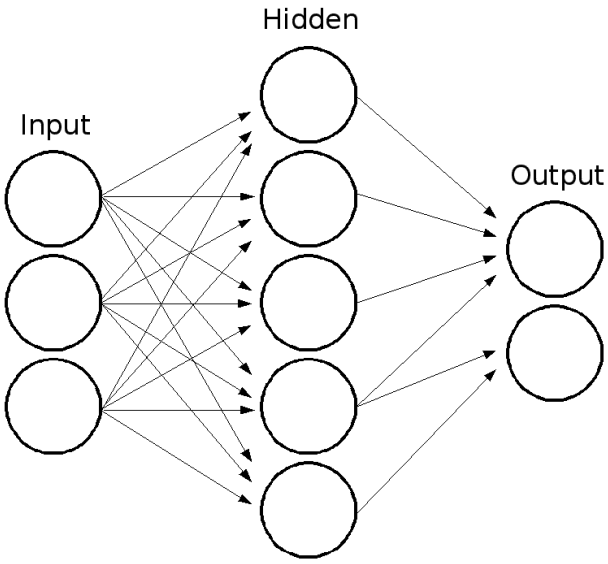


Figure 1. A schematic representation of an artificial neural network.

Image derived under the GNU Free Documentation License from https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg, by user Cburnett.

At its most basic, an artificial neural network consists of an (often large) number of interconnected simple units, or nodes. The nodes are analogous to the brain's neurons, and the connections between them are analogous to synapses. Nodes are arranged in layers: an input layer, which receives the information to be processed, an output layer, and zero or more intermediate layers, known as hidden layers.

[156] McCulloch and Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity."

Each node has an *activation value*. As nodes receive signals from the layer beneath them, their activation values are modified (increased or decreased). If a node's activation value exceeds its threshold, the node will fire, sending the signal forward to the next layer.

The *weight* of the connection between any two nodes modifies the strength of the transmitted signal. The strength of that signal then affects the activation value of the receiving node. Connection weights can be thought of as the diameter of the 'pipe' between two nodes.

Connection weights can be modified, with the result that a given input at the bottom layer results in a different output at the top layer. This means the neural network can learn — if the output of the network is what is wanted, the connection weights are increased. If not, the weights are decreased. Through a process of iterative training (trial-and-error), often involving thousands of input samples, the connection weights gradually approach a configuration that will reliably respond in the desired way.

For example, an artificial neural network may be trained on numerous sample images of cats. After training, the network will be able to detect cats in previously unseen images.

Parallel Processing

Connectionist researchers modelled their artificial neural networks on the parallel processing structures found in nature. Animal brains divide complex tasks into a number of sub-tasks, which run in parallel. For example, vision is divided into shape, colour, depth and motion components. Each of these components is analysed separately in parallel, before being recombined into the visual field we experience.

> The neural networks found in slugs, hamsters, monkeys, and humans are … vast parallel networks of richly interconnected, but relatively slow and simple, processors. The relative slowness of the individual processors is offset by having them work in a cooperative parallelism on the task at hand.[157]

[157] Clark, *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*, 4–5.

When an artificial neural network is supplied with input data (such as an image), the activated nodes in the network's input layer fire simultaneously, causing a large number of parallel signal cascades to flow through the network.

Parallel Distributed Processing (parallel processing using *distributed representations*) is the dominant form of Connectionism and the terms are often used interchangeably.[158]

Bottom-Up and Top-Down Interaction in Neural Networks

The above description corresponds to the most basic form of neural network, a *feedforward* network. The flow of information is one way — from the bottom (input) layer, to the top (output) layer. Feedforward networks are therefore an example of the bottom-up information processing paradigm. More complex networks have top-down feedback connections in addition to feedforward connections. These are known as *recurrent neural networks*.

In both biological and artificial perceptual systems, top-down knowledge provides additional functionality, such as the ability to disambiguate fragmentary data sources. For example, during an optical character recognition task, a neural network can refer to stored representations of letterforms to disambiguate noisy or damaged input images. The top-down signal is fed back through the same path as the bottom-up signal, and modulates the bottom-up signal (by modifying firing thresholds for example).[159]

Two Consequences of Distributed Representation

Information in a neural network is represented by the pattern of activity across the entire population of nodes and the connections between them. The *activity pattern* is a complex pattern of connection weights and firing thresholds across a large number of nodes. This differs fundamentally from symbolic systems, in which representations are composed of discrete, word-like units — the symbols.

[158] A distributed representation is the pattern of activity across a neural network in response to a given input.
[159] Fukushima, "Artificial Vision by Multi-Layered Neural Networks," 114-117.

One consequence of distributed representation is that the representations carry semantic meaning. Andy Clark discusses a neural net trained to perform optical character recognition.[160] Since the network responds to the characteristic similarities and differences between the shapes of letters, its activity pattern (internal representation) for the character E will be more similar to an F than it is to a U. Similarly, a network's activity pattern for horse images will be more similar to the activity pattern for donkey images than to house images. Compare this to the linguistic representations of 'horse', 'donkey' and 'house'. The aural and visual properties of 'horse' and 'house' are more similar than those of 'horse' and 'donkey'. This is simply the fortuitous consequence of the arbitrary relationship between words and their referents. In contrast, the activity patterns in which distributed representations consist are not arbitrary, but embody semantic information.

> The semantic (broadly understood) similarity between representational contents is echoed as a similarity between representational vehicles. Within such a scheme, the representation of individual items is nonarbitrary.[161]

This also means neural networks are able to generalise, to respond appropriately to cases which do not precisely fit the characteristics of their training set. The optical character recogniser above can respond to an image that looks more like an E than it looks like any other letter.

A second consequence of distributed representation is opacity. It is difficult or impossible for an observer to understand how the complex patterns of connection weights and firing thresholds in a neural network relate to the input data set. The connection weights reflect the complex interactions of many factors, some of which may not be known to the observer. For example, analysis of NETtalk (a 1980s text-to-speech network) "reveals that the net learned to represent such categories as consonants and vowels, not by creating one unit active for consonants and another for vowels, but rather in developing two different characteristic patterns of activity across all the hidden units".[162]

Thus, distributed representations are not human-readable.

---

[160] Clark, *Associative Engines,* 19.
[161] Ibid., 19.
[162] Garson, "Connectionism."

Learning in Neural Networks

As artificial neural networks demonstrate, bottom-up processes using distributed representation are able to learn, and do so constantly by modifying their structure in response to experience. They learn to recognise patterns in input data. Those patterns are the network's concepts.

As Geoffrey Hinton, one of the pioneers of Connectionism and Machine Learning puts it:

> You have to learn to recognize things without anybody telling you what the things are. Then after you learn the categories, people tell you the names of these categories. So kids learn about dogs and cats and then they learn that dogs are called "dogs" and cats are called "cats".[163]

In contrast, top-down symbolic systems such as today's personal computers perform computations on pre-existing knowledge. In a symbolic computer system, the meaning of the symbols ultimately reside in the head(s) of its programmers and users. The computer manipulates the symbols on the basis of their shape (syntax), not their meaning (semantics). Such a system can prove or disprove hypotheses based on existing concepts, but it cannot acquire foundational concepts. This is a consequence of the conventional meaning of symbols.[164]

Today's artificial neural networks are highly simplified models of biological brain structures. Their performance is not comparable to human performance in those tasks we have evolved to excel at (such as face recognition and spoken language). Their performance is human-comparable in certain perceptual tasks (for example visual identification of traffic signs and mitosis), and they are capable of detecting novel patterns in other data formats which humans are not adept at reading (for example, streaming data from sensors in technical equipment).[165]

[163] Wolchover, "As Machines Get Smarter, Evidence They Learn Like Us."
[164] Horst, "Symbols and Computation: A Critique of the Computational Theory of Mind."
[165] Schmidhuber, "Deep Learning in Neural Networks," 97–99; Veta et al., "Assessment of Algorithms for Mitosis Detection in Breast Cancer Histopathology Images."

Author/s:
Lines, Jeremy

Title:
Voiceover narration and audio-visual imagery in non-fiction film

Date:
2015

Persistent Link:
http://hdl.handle.net/11343/91084