Inspiration and Novelty:

Creativity and the Modularity of Mind

Andrew McIntyre

Oxford Brookes University

Submitted in partial fulfilment of the requirements of the award of

Doctor of Philosophy

September 2014

Inspiration and Novelty:
Creativity and the Modularity of Mind

Abstract

Creativity is the channel by which new ideas are generated, and through which all new and valuable things of any kind, artefacts or ideas, are added to human culture. As the entire cultural environment is composed of entities which are either themselves new or are copies of things which once were new, the subject is of some ontological significance. It is therefore regrettable that it is still far from completely understood, even though there has been considerable progress in recent years. The thesis examines the current state of philosophical thinking on the subject, and also on modularity of mind, which it takes as a framework against which to investigate creativity. It then sets out to work towards a naturalistic explanation of the origin of novelty in creative thought. Darwinian evolution by natural selection explains the origin of novelty in the biological sphere, and theories of creativity drawing an analogy with evolution are examined. It is concluded that whereas creativity clearly shares some of the most general characteristics of evolution, there are crucial points of variance. The outcome of the investigation draws particular attention to the known requirement for the creative product, which does not have a direct evolutionary analogue. Focusing on the requirement and the direction of internal and external searches is shown to provide for the inception and exploration of possible creative solutions and also to circumvent the associated *Meno* paradox. This in turn is the basis for the positive assessment of the feasibility of proceeding to model a creative process. The study culminates in the construction of such a model, at a necessarily high level of aggregation. This is offered as a candidate route towards a naturalistic explanation of the genesis of new ideas and cultural innovation.

Statement of Objectives

Aims

The investigation was set up with the following initial aims:

- To investigate the nature of creativity, in the light of recent thinking on modularity and evolutionary psychology.
- More particularly, to investigate whether the massive modularity thesis has the potential to support a defensible theory of creativity, offering an explanation as to how human creativity is possible, and how it relates to other aspects of mentation.

Objectives

The aims were broken down into specific objectives:

- To conduct a search of the relevant literature on creativity, modularity, and massive modularity
- To subject the identified literature to critical review and analysis
- To confirm the feasibility of modelling a creative process
- Subject to the outcome of the feasibility assessment, to carry out a modelling exercise as above.

Expected outcomes

The expected outcomes and outputs from the study were therefore:

- A critical review and analysis of the literature
- An assessment of the feasibility of modelling a creative process; and
- A model of a creative process (dependent upon the feasibility assessment).

Acknowledgements

I would like to record my gratitude for the assistance received from the supervisory team, and especially Mark Cain and Constantine Sandis, at the Department of History, Philosophy and Religion of Oxford Brookes University. I am also indebted to Maria Kronfeldner of Bielefeld University, for kindly supplying a copy of the manuscript for *Darwinian Creativity and Memetics*.

Contents

	Page
Preface: the background to the research	9
Thesis structure	11
Part One: Foundations	
1. Chapter one: *Creativity and the Origin of Novelty*	13
1.1 Introduction to the first chapter	13
1.2 Creativity: a few starting points	16
1.2.1 The scope of creativity	18
1.3 Questions about creativity and the origin of creative novelty	19
1.3.1 The mysterious nature of creativity	19
1.3.2 The *Meno* paradox	20
1.3.3 Creative novelty and spontaneity	22
1.3.4 Creativity and the origination of novelty	23
1.3.5 Creative novelty and inspiration, or intervention	25
1.4 Towards an account of creative novelty	26
1.4.1 Inspirational, or interventionist, theories	27
1.4.2 Creativity and subliminal mental activity	28
1.4.3 Theories favouring sources in pre-existing mental content	28
1.4.4 The sources of creative novelty	30
1.4.5 Darwinian theories and the proposed Darwinian analogy	32
1.4.6 The role of environmental challenge	33
1.4.7 Searching and the *Meno*	35
1.4.8 Mental architecture	36
1.5 Evolutionary psychology	37
1.6 Language and creative thought	41
1.6.1 Metaphor and analogy	44
1.7 The four stage process model	45
1.8 Summary and conclusion	47
1.9 Examples of introspective reports of creative insight and serendipitous discovery	48
2. Chapter two: *Creativity and Context*	51
2.1 Introduction	51
2.2 What is ‘creativity’?	56
2.2.1 Novelty and value and their relation to context	57
2.3 Novelty	.63
2.3.1 Nature and Art, or culture	63
2.3.2 Possible and actual	67
2.4 Varieties of novelty and their relation to creativity	71
2.4.1 Anthropological novelty and anthropological creativity	71
2.4.2 Historical and psychological varieties of novelty	72
2.5 Varieties of creativity	75
2.5.1 Historical and psychological varieties of creativity	77
2.5.2 Metaphysical novelty and creativity	75
2.5.3 Causation, determinism, and *ex nihilo* creativity	78
2.5.4 Varieties of creativity relative to surprisingness	81
2.6 Potential sources for creative ideas	85

2.7 Creative process 90
2.7.1 Possibility and actuality 91
2.7.2 The Universal, or Total, Library of Babel 92
2.8 The *Meno* problem 97
2.8.1 Proposed modifications to the evolution by natural selection analogy 101
2.9 Conclusion 102

3. Chapter three: *Modularity* 104

3.1 Introduction 104
3.1.1 Objective: the potential benefits of a modular framework 106
3.2 Modularity according to Fodor 107
3.2.1 Fodorian mental modules 107
3.2.2 Fodor's central systems 108
3.3 Discussion: Fodorian modularity 113
3.4 Massive modularity 120
3.5 Carruthers and *The Architecture of the Mind* 120
3.6 Discussion: Carruthers and massive modularity 125
3.7 Carruthers' implied process model. 129
3.8 Summary of main points 130

Midway: Linking Foundations to Modelling

4. Chapter four: *Formulating a Candidate Creative Process* 132

Part Two: The Creative Process

5. Chapter five: *Feasibility and the Darwinian analogy* 150

5.1 Introduction 150
5.2 Aims for modelling 153
5.2.1 Methodology 154
5.2.2 Benchmarks 155
5.2.3 Practical, realistically attainable expectations for outcome 156
5.2.4 The basis for an assessment of feasibility 157
5.3 Stepping through a creative process 158
5.3.1 Evolution by natural selection as an analogy for human creativity 160
5.4 The proposed Darwinian analogy 166
5.4.1 Kronfeldner's analysis 169
5.4.2 Discussion 174
5.5 Conclusion to chapter five 190
5.5.1 Methodological issues 191
5.5.2 The Darwinian analogy 191
5.5.3 Sources of novelty 192

6. Chapter six: *Modelling a Creative Process* 194

6.1 Introduction 194
6.1.1 Coverage 196
6.2 Model foundations 196
6.3 Preliminary sketches for modelling 198
6.3.1 Modularity 198
6.3.2 Staging 199
6.3.3 Darwinian evolution 202

6.3.4 Prior knowledge and the requirement 204
6.3.5 Search and query 206
6.3.6 Matching and recognition 210
6.3.7 Evaluation 211
6.4 Constructing the complete high-level model 213
6.5 Modularisation 215
6.6 Counterarguments 220
6.7 Conclusion 227

7. Chapter seven: *Findings* 228

7.1 Introduction 228
7.1.1 Coverage 228
7.2 The proposed creative process and the criteria for success. 229
7.2.1 Acquisition of prior knowledge and the formulation of the requirement 230
7.2.2 Searching 231
7.2.3 Searching within and across domains of thought 233
7.2.4 Matching and recognition 235
7.2.5 Evaluation 235
7.3 The model as a process 235
7.4 The model considered as explanation 236
7.4.1 The *Meno* paradox 237
7.5 Overall assessment of the outcome of the modelling exercise 238
7.6.1 Unresolved issues 239
7.6 Conclusion 239

8. Chapter eight: *Conclusions* 240

8.1 Introduction 240
8.1.1 The course of the project 241
8.1.2 Outcomes 245
8.2 Summary of conclusions and implications 246
8.3 The modelled creative process compared against natural selection, by stage 250
8.4 The Darwinian analogy 256
8.5 Is the modelled creative process original? 259
8.6 Modularity 262
8.7 Results and findings 264
8.8 Assumptions and risks 267
8.9 Implications 270
8.9.1 The mystery of creativity 270
8.9.2 Designers, virtual testing, and the environmental drivers of creativity 271
8.9.3 Automation 273
8.10 Further directions and opportunities created by, or raised during the course of, the study. 274
8.10.1 And finally… 275

Bibliography 277

List of figures

2.1: Mill's Nature and Art 65
2.2: Creative process, relative to Mill's Nature and Art 66
2.3: Creative process, relative to possible and actual 68
2.4: Varieties of cultural novelty (macrocosmic viewpoint) 73
2.5: Varieties of cultural novelty (microcosmic viewpoint) 74
2.6: Evolution by natural selection: inputs and outputs 87
2.7: The creative generation of cultural novelty – potential input sources 89
3.1: Fodorian input module 109
3.2: Quasi-modular central systems process (conjectured) 117
5.1: Two-phase selection of cultural novelty 176
5.2: Culture and nature 177
5.3: Natural selection (after Darwin); viewed as a two-stage process 179
5.4: Copying, learning, and creative output 185
5.5: Spontaneity and routine production. 186
6.1: Classic Fodorian input module (encapsulated, domain-specific) 198
6.2: Wallas stages (annotated) 199
6.3: Extended staging 200
6.4: Natural selection (generalised) 202
6.5: Extended staging (with mapping to natural selection) 203
6.6: Generation of the requirement 206
6.7: Query and search 207
6.8: The search box 209
6.9: Recognition and evaluation 211
6.10: Complete process model 214
6.11: Fodorian input module and search and query module compared 217
6.12: Search and matching/recognition modular framework 218
6.13: Core sub-Fodorian modules 220
6.14: Possibility space and conceptual space 225
7.1: Proposed metastructure of domains 234
8.1: Evolution by natural selection (from fig. 2.6) 250
8.2: Natural variation and selection (from fig. 6.4) 251
8.3: Complete process model (from fig. 6.10) 252
8.4: Classic Fodorian input module (encapsulated, domain-specific), from fig. 6.1 263

Preface: the background to the research

The thesis reports the outcome of a research project investigating creativity, with particular emphasis on the central mystery of the genesis of novelty, the process by which new ideas come to be conceived.

Creativity is the pathway through which original thought is first brought into being. All constituents of human culture are the embodiments of either novel ideas, or copies of earlier notions that were new at some time in the past. Creativity is, therefore, of considerable significance as the ground of the entire artificial environment. It is intimately connected with newness, or novelty. It also envelops the underlying problem of origination: how can new things come about in a world where energy is conserved but not created, and something cannot be obtained for nothing? Any wholly or partly creative act adds a contribution to the cosmos that simply was not there before; it is an ontologically significant event. It takes what was previously only possible (and unknown), and renders it actual.

Insofar as it gives rise to the entirely new, creativity participates in the greater question of existence, as to how anything *is* at all. The novelty resultant from creativity is only a part, a subset, of a much larger category, because all the constituents of the cosmos, including the material universe itself, elements, stars, planets, and life, were themselves once new. Creativity is one of the channels through which novelty is mediated; it generates new properties and added value, and thereby has something in common with biological evolution, which is likewise a source of novelty. In creativity, as in evolution, more is output than was there beforehand: novelty and value result. In the case of biological nature, in modernity the phenomenon is explained by natural selection: individuals vary, more successful variants are favoured by environmental pressure, and these characteristics are transmitted and preserved. All it takes is the operation of natural law, nothing more; no designers are called for. But with human

artefacts and innovation the generation of novelty seems a degree more puzzling. Humans are designers and have new ideas, and so where do those new ideas come from? This is the principal question addressed by the thesis. The origination of creative thought is particularly obscure, as many have noted: ideas certainly can give the impression of having sprung out of nowhere or nothing. They may surface in dreams, or while the creative individual is engaged on something other than the problem in question.

It is a basic assumption of this thesis that creative processes can be investigated in the reasonable expectation that explanation is achievable, and in recent years a number of philosophers and cognitive scientists have turned their attention to the issue, and considerable progress has been made in that direction. Modern philosophical opinion on mind has generally favoured a materialist line, to which creativity (like consciousness) presents a challenge, but theories favouring modular architectures offer promising new lines of approach towards explaining mental process and function, including creative thought.

This research project has aspired to set up a modular framework to support the construction of a high-level model for a creative process and thereby contribute towards its explanation. Darwinian evolution by natural selection establishes a process which explains the origin of biological novelty but there is no equivalent, generally accepted, model for the origin of cultural novelty. The current project has taken evolution by natural selection as a candidate analogy for creativity, as have others, but its distinctive approach has been to identify and emphasise the points of divergence between the two, as well as the correspondences. This has been the key to the derivation of the model, which is based upon an initial understanding of the required properties, which can then be used to focus a search for their embodiment in a creative product.

Creativity is often thought of as a distinctively human capacity, although there is no *prima facie* reason to suppose that creative products might not conceivably be generated by artificial systems, nor that there is only one possible process which might be capable of creative output. Hence a successful model, capable of creative output, may or may not represent human creative process, or the only possible creative process, but such a model would suggest a route by which creative novelty could be generated, which would itself represent a significant advance.

The modelling exercise was preceded by an intensive exploration of the relevant literature, which raised further issues to be taken into account in setting up the model, and these in turn provoked the development of assumptions, principles, and steps which subsequently informed the modelling.

The construction of the model, and the prior investigation of the literature, are covered in the body of the thesis as outlined below.

Thesis structure

In brief, therefore, the focus of the research is on the central mystery of creativity, the genesis of creative novelty. At present there is no known (or at least, no generally acknowledged) *process*, no set of logical steps or stages, that lead from any antecedent state to the origination of a novel, creative, thought or product. This contrasts, for example, with the steps understood to lead from premises to conclusion in deductive logic, or the sequence of arithmetical operations involved in calculating the circumference of a circle, given its radius.

Creativity and creative novelty are by no means simple and straightforward notions, and I have accordingly devoted the first two chapters to an exploration of the general scope and nature of creativity and questions surrounding the origination of creative novelty. The third is assigned to modularity, which will form the architectural background for the model. In the fifth chapter I proceed to gauge the feasibility of

constructing a model for a creative process, informed by the results of the earlier investigations. After considering the counterarguments, I opt for a model based on a modified Darwinian analogue, the construction of which is detailed in chapter six. In the seventh chapter I evaluate the outcome of the modelling exercise, and the overall conclusions from the study follow in the final chapter.

Generally speaking, the structure of the thesis conforms approximately to the pattern set by Wallas's four stages (as described in chapter one). The first three chapters record the acquisition and evaluation of prior information on the relevant fields, including: creativity, novelty, modularity and Darwinism. The fifth chapter focuses on issues having a particularly direct bearing on the construction of a model, and might stand comparison with incubation. The sixth chapter covers the modelling exercise itself, and therefore the development of the modelled creative process. The final chapters contain the verification and assessment of the potential impact of the outcome of the project.

PART ONE: FOUNDATIONS

Chapter One: *Creativity and the Origin of Novelty*

1.1 Introduction to the first chapter

In this study, I am seeking an explanation for the origination of creative novelty, in terms of a plausible process that will conceivably be capable of generating output satisfying appropriate criteria for creative products. The focus therefore is very much on *how* creativity can happen, rather than *what* creativity is. But, it is clearly necessary to understand as precisely as possible the nature and characteristics of creativity in order to determine what it is that any such process will have to mediate. Creativity is widespread throughout human culture and the artificial environment, and there is no universally accepted definition. Also, there is always the chance that creativity is not a homogeneous concept, with more than one phenomenon, or generative process, being involved. It might be necessary to restrict the scope of any constructed process to cover only a part of what is generally understood by creativity, in order to provide a feasible brief for the modelling exercise which is to follow.

This chapter sets the scene for the research project and for the main body of the thesis. It explores some of the principal issues surrounding the study of creativity. The main problems with which this project has been concerned are set out, and the approach to be taken to them is outlined. The core question addressed by the investigation is itself tightly demarcated: to enquire whether a model can be constructed which will represent, or correspond to, a process capable of generating creative output. But, it is positioned within what is a very complex field. Not only does creativity have its own mysterious aspects but it also touches on other, less than

straightforward, issues. This chapter surveys the broader picture, with the aim of isolating and extracting the main points having a bearing on the central theme of the mental origin of cultural novelty. Where appropriate, these will inform the more detailed consideration in the later chapters, the distilled outcomes of which will ultimately be gathered to focus on the modelling exercise.

Creativity cuts across subject boundaries, as it is found in all areas of human activity: science, humanities, technology, office administration, medicine, home maintenance, and anywhere else we might care to think of. Hence it is necessarily a cross-discipline topic and can only be treated as such. This is an indication of its importance and its key place and presence within the entire field of human culture. All this, together with its general ontological significance, and compounded by its still relatively underresearched status (which this cross-sector aspect may help to explain), suggests at once both the scale of opportunity for further development, and also the dangers accompanying such a broad scope. Maria Kronfeldner (2007, p.iii) makes a similar acknowledgement of risk, but there is no avoiding it except at the cost of ignoring such a clearly significant issue[1]. Creativity touches on many other areas of study, which often have their own (considerable) points of uncertainty, so that decisions have to be made as to the lines to be taken on them, as it would be impossible, from their sheer numbers and complexity, to go into much detail. I will return to this in the final chapter.

Accordingly, I begin with a wide-ranging survey, or exploratory chapter, taking an overview of creativity and some of the areas neighbouring or linked thereto, in order to achieve a broad acquaintance with the field; I will follow this in the second chapter with a more concentrated investigation of creativity, and also novelty, in detail. The chapter starts with a general examination of the full scope of the topic, leading into an initial scrutiny of the nature of creativity, the key novelty and value criteria, and the reasons for the cloak of mystery which is often thought to surround it. The element of mystery

[1] See also chapter eight.

particularly attaches to the origin of novelty, the aspect of central interest and relevance to the current project. Part of it derives from the very nature of novelty and origination, and perhaps also of the process by which creative novelty is generated. If an idea is new, by definition it was unknown beforehand, so it almost necessarily is accompanied by an element of the unexpected.

I then proceed to cover more specific questions, which any attempt to account for the generation of creative novelty will need to address, or at least acknowledge. These include: the relation of the new, original mental content to anything that came before, the obscurity of any intervening processing, and the difficulty of searching for the unknown (as in the *Meno* paradox).

Next, I indicate a few of the lines that various theorists have taken on some of the problems previously identified. These include the generation of novelty through inspiration by an external agency or muse and the incubation of creative thought in states of restricted awareness, or in the absence of the reasoning faculty. Others have sought to draw an analogy with the operation of organic evolution through natural selection. Darwinism offers a candidate analogue because, in common with creativity, it deals with the origin and generation of novelty. The treatment in the current chapter is confined to a brief introduction, to position the topic relative to other general issues. This leads into a preliminary sketch of aspects of the approach I intend to adopt and flesh out in the later chapters, which calls for the development of a modified Darwinian scheme.

Darwinian arguments are also employed in evolutionary psychology in support of explanations for the development of mental function, particularly in the context of modularity theories. Modularity is to be taken as a framework for modelling, and hence an example of this type of argument is discussed and examined here. Evolutionary psychology is of course of particular interest in relation to the development of specifically human creativity.

I then move on to commence investigation of the kind of process that has been thought to be involved in mediating the origin of creative novelty. I consider the four-stage process model introduced by Wallas in the early twentieth century, and since widely adopted. This will later be taken as one of the starting points leading up to the development of the modelled process in chapter six.

There follows a summary of the main points from the chapter, to which I add a brief outline of a few examples of introspective reports of creative thinking and of serendipitous discovery.

This first chapter is thus intended partly to give an idea of exactly how broad the potential coverage of the topic is; also to introduce the main problems surrounding creativity which are to be addressed and the principal lines of argument which will be pursued. As a result, many questions are raised at this stage for which resolution, or further discussion, will be deferred to later chapters, and therefore the full significance of some areas may not be immediately apparent.

1.2 Creativity: a few starting points

In its most general sense, ‘creativity’ refers to bringing something into existence. This initial description already provokes questions as to ontological status and significance: just how does it happen that ideas and types of artefact that had not been there before, and indeed were completely unknown, come to be? And this problem of origination is not the only puzzling aspect associated with creativity, as will be discussed below. Yet much of the environment we live in derives from human creative endeavour, whether directly or as a result of the replication of other things or ideas which ultimately trace back to an origin in human creative activity. So, creativity is both widespread and incompletely understood; hence one of the drivers behind the current interest.

To begin with, there are conditions attached to the thing that is brought into being if it is to qualify as creative output, as alluded to in Noam Chomsky’s *Cartesian Linguistics*:

> For Humboldt, as for his Cartesian and romantic precursors, the normal use of language typically involves creative mental acts; but it is the character of a language rather than its form that reflects true “creativity” in a higher sense – in the sense that implies value as well as novelty. (2009, p.75).

The end result does have to be new (to have had no prior existence) but this is not sufficient by itself, as not all new things are regarded as evidence of creativity. Uttering nonsense, or the compilation of meaningless sentences such as: ‘*Tomorrow’s newspaper used to churn tomatoes within a melodious brick*’ is not generally considered to be creative, even if the output is grammatically correct and completely original (that specific string of words never having been put together before). This is because the novelty is unaccompanied by any added value. Margaret Boden provides a similar example in *The Creative Mind* (2004, p.13) but in that case the position alters when the nonsense saying is taken in its full context, because it is an integral part of Boden’s book, itself very much a product of creative thought and action. The reason for this shift of classification is that the nonsense does add value when viewed against its setting. Novelty and value are both context-sensitive, as will be explored in chapter two, where more will be said on the nature of creativity, although for now novelty and added value can be taken to characterise it.

Similarly, reproducing a Shakespeare sonnet, even if there is a use for the copy (so that additional value is generated in the absence of novelty), is not considered to be a creative act.

In sum, the two most basic criteria for creativity are: novelty and value. Neither is sufficient on its own and arguably even taken together they are not restrictive enough. These properties give rise to a number of other issues (see chapter two), but for present purposes it will be adequate to note that creativity is how we get new entities and properties which are of value, in human culture. Indeed, culture is made up of products of creativity, both original (as in Shakespeare writing *Macbeth*), and as subsequently copied and modified (and performed, in *Macbeth*’s case).

1.2.1 The scope of creativity

Taking all human activity generating novelty and value to be creative gives a broad reach: 'creativity' is here taken in an inclusive sense, so as to incorporate mundane examples such as the development of a new refinement to an office procedure. The scope does encompass especially significant creative acts, including major scientific discoveries and the composition of great music, but is by no means limited to these. Nor is this study restricted to the work of particularly talented creative individuals.

Even though creativity involves a (human) agent, novelty and value both refer solely to the output of a creative process, the end product or result. There are nevertheless other entities required for any creative act, and these do include both the creative individual, or agent, and the process by which the output is generated, as Albert Rothenberg and Carl Hausman note:

> Minimally, … creativity consists of the capacity for, or state of, bringing something into being. And bringing something into being involves at least three separable components: an agent, a process, and a product. (1976, p.6).

Definitions often focus on the properties of the output as the primary determining factor in identifying instances of creativity. A creative product in the end has to be presented in some physical form – unless it is to be left as an unexpressed thought – so the output is potentially publicly observable. (It must also be empirically possible, even if the ideas represented are not, as will be explored in chapter two.) In contrast, the roles of the agent, and of the process by means of which the creative output is brought forth, can be relatively inaccessible and opaque, as discussed below.

To the three above-mentioned contributing entities (producer, product, and process), I would add a fourth: input in terms of previously known elements, constituents, or ideas. For example, Shakespeare drew on various sources, including, when writing *Macbeth*, Hollinshed's *Chronicles* (Muir, 1984, p.xxxvi). This component, the source, in which I would include any contributions from precursory thoughts, concepts, images,

perceptions, or other mental content, is more controversial than the other three because some views of creativity allow for inspirational sources (or none at all), which are presumably potentially independent of antecedent conditions. It therefore touches on the central question of the origin of creative novelty, which is at the heart of the matter under investigation.

1.3 Questions about creativity and the origin of creative novelty

This section introduces the main questions that will be discussed in more depth in later chapters. The central problems concern the origination of novelty and the relation of original thought to antecedent mentation, or to other forms of input, that is to what I have termed 'the source' (above). Some of the difficulty bears simply on the nature of novelty in its widest sense, and the issue as to how the new relates to what was there before. There is also the similar psychological question of the relation (or absence of relation) of new, original thought to previously held knowledge or belief; this arises partly because of the opacity of any mental processing leading to creative output. The *Meno* paradox is also included, because of its relevance to the problem of searching for an unknown.

I begin with a general outline of the mystery surrounding creativity, novelty, and creative novelty in particular.

1.3.1 The mysterious nature of creativity

There is a multi-faceted element of mystery surrounding creativity, some aspects of which have already been mentioned. As said, one advantage of taking novelty and value as criteria for creativity is just that they apply to (empirically possible and observable) creative output, whereas the mysterious side of creativity attaches more to the role of the agent and to the creative process. The process through which a creative product is brought forth is much more resistant to inquiry. It is generally impenetrable to introspection: creative individuals often cannot say much about where their ideas

come from. Boden (in *The Creative Mind*) neatly encapsulates their accounts in the phrase: '*The bath, the bed, and the bus'* (2004, p.25). '*The bath*' (Archimedes), and '*the bus*' (Poincaré) illustrate the way in which a eureka moment can occur when the individual is not actively engaged on the problem concerned, but rather as an insight whose immediate precursors in thought are not readily discernible. Put another way, creative thought can appear to come from nowhere, or perhaps out of nothing. Similar considerations apply to serendipitous discovery, as in the case of Fleming (see section 1.9 below).

1.3.2 The Meno *paradox*

It is fundamental to the nature of creativity that creative solutions are not accessible in advance: as soon as they become known, they are no longer new and the initial phase of creativity has passed. Linked with this is the question as to how originality is indeed even possible at all, because if the solution is unknowable beforehand, how then can it be sought when there is no indication as to what to look for, or how or where to find it? This paradox features in Plato's *Meno* (in which Socrates elicits the solution to a geometric problem from an untutored slave)[2].

> MENO. But how will you look for something when you don't in the least know what it is? How on earth are you going to set up something you don't know as the object of your search? To put it another way, even if you come right up against it, how will you know that what you have found is the thing you didn't know?
>
> SOCRATES. I know what you mean. Do you realize that what you are bringing up is the trick argument that a man cannot try to discover either what he knows or what he does not know? He would not seek what he knows, for since he knows it there is no need of the inquiry, nor what he does not know, for in that case he does not even know what he is to look for. (1956, 80D-80E, pp.128-129).

Socrates is of the view that the new information is obtained through recollection of before-life experience (see: 81C-81D, 1956, pp.129-130). And indeed a flash of creative insight might be taken to bear some resemblance to memory recall, in that

[2] See Nickles (2003, p.56), and also chapter two, below.

remembering can also be spontaneous, and its process similarly inaccessible to introspection.

The *Meno* paradox also raises the question as to how creative novelty might be recognised, given that it has been found. As noted above, it can even be queried on this basis as to whether original thought can actually happen, as Thomas Nickles points out (2003, p.56). The paradox is discussed further in chapter two.

1.3.2 Subliminal processing?

Another reason why the creative process can be obscure relates to Boden's reference to '*the bed*' (above), in that it might be thought to occur below the threshold of consciousness. And indeed there are frequent references to sleep and dreaming (or day-dreaming, in Kekulé's case – see later on) and other states of restricted or altered awareness, in the accounts of creative individuals. This prompts the question: is there something about sleep that, besides '[knitting] up the ravell'd sleave of care'[3] as Macbeth (Shakespeare, 1984, 2.2:36) has it, in some way facilitates creative thought? It also reinforces the opacity of the process by which human creativity is mediated, or indeed of any process that might result in creative output, as another aspect of the mystery surrounding creativity. Associated with this is the observation that creativity is not (or at least not directly) amenable to the will. Immanuel Kant noted both the involuntary aspect and the obscurity of any antecedent mentation:

> … Genius itself cannot describe or indicate scientifically how it brings about its products, and it is rather as *nature* that it gives the rule. That is why, if an author owes a product to his genius, he himself does not know how he came by the ideas for it; nor is it in his power [*Gewalt*] to devise such products at his pleasure, or by following a plan, and to communicate [his procedure] to others in precepts that would enable them to bring about like products. (Indeed, that is presumably why the word genius is derived from [Latin] *genius*, [which means] the guardian and guiding spirit that each person is given as his own at birth,[…] and to whose inspiration [*Eingebung*] those original ideas are due.) (1793, Ak.308, p.175, footnote omitted).

[3] Boden refers to this line in another, slightly different, context (2004, p.2).

This aspect, taken together with the *Meno* problem, bears on the problematic character of creative process, and the question as to how creative novelty is, or can be, generated. Central to the issue is the relation of creative novelty to prior knowledge and antecedent conditions, and the nature of the source material (if any) from which the creatively novel notion is to be formed.

1.3.3 Creative novelty and spontaneity

Creative products have to include an element of novelty, and so they do not seem to arise from the application of known or routine algorithmic procedures, such as the calculation of an arithmetical expression, nor from logical, deductive reasoning. Here, the solution might be regarded as implicit given a known procedure with specified inputs or starting conditions, so that in such cases, the outcome could be seen as having been accessible in advance, and creative novelty is compromised. The novelty criterion thus rules out standard problem-solving procedures. Creativity requires that the output be unknown beforehand (otherwise it is not creative, because no novelty is originated). Novel (and, to an extent, unexpected) properties or features must be produced or generated, so that they almost necessarily appear to be spontaneous and/or surprising, or perhaps to spring out of nowhere or nothing.

If this observation is turned around, it becomes the question as to how any explanation of creativity might be achievable, on the assumption that it has to be explained as the effect of necessary and sufficient causes. Once the causes are known, it might be assumed that the effect is anticipated in them (in the same way that a syllogistic conclusion might be considered to be anticipated in the premises from which it is derived)[4]. This leads to a paradox (not dissimilar to the *Meno* paradox, or the ones that Kronfeldner (2009, p.582) and Hausman (1984, pp.9-10) describe – see also chapter two. If there is no external, supernatural input, creativity should be explainable in terms

[4] See also Hausman (1984, p. ix).

of its causes. But then if creativity demands novelty, how can it be possible unless the creative output could *not* be anticipated from antecedent causes?

These issues will be discussed further in chapter two. But for the time being, it can be noted that one aspect of the mystery of creativity is that creative insight cannot be anticipated in advance, as this is the other side of the requirement for novelty. Any creative process is thus in some sense almost necessarily opaque.

1.3.4 Creativity and the origination of novelty

A new poem is not constituted of, nor explained by, the collection of letters, words and punctuation through which it is expressed. The novel creative output does seem to add something, effectively from nothing, which overlays the combinations of symbols; otherwise all random arrangements of the same collection of symbols will be more or less alike, and generally speaking would not be expected to form additional new poems. Thus at least purely in terms of the separate constituents (if not in terms of effort, previously acquired experience and so on) the creative product seems to add something from nothing – and maybe in a sense, *for* nothing – as the new poem is not a property of the symbols by means of which it is expressed in writing.

The position becomes more complex when we consider the similarity of the appearance of new things, and the concept of novelty, with that of the cosmic mystery of existence itself[5]. And indeed the nature of the mystery is to an extent responsive as to whether creativity is viewed as an internal process, or as a cosmic event (as bringing something new into the world); in some respects the element of mystery relates more directly to the nature of what is occurring relative to the external universe when creativity is encountered. This is the ontological question of innovation referred to above. It is simply the problem as to how new ideas, things, or properties, can arise at all. In a creative process, output might be produced that is cosmically novel; it is in this way that all the features of human culture originated, none of which were known prior

[5] See for example Kaufman (2004, pp.55-56).

to their invention or discovery (and human culture is also a repository of immense value: the stored results of the past generation of new ideas, and the subsequent copying and modifying of those found to be worthwhile).

In Alfred North Whitehead's 'Speculative Philosophy' (1985, p.3), creativity is viewed as 'ultimate' (1985, p.7) and as occupying a key position relative to the universe as a whole. Whitehead's creativity is associated with pure novelty (1985, p.21) and is not limited to evolution by natural selection and/or human creativity (which requires the generation of value as well as novelty). Is human creativity, then, a subset of, or does it (as a key to unlocking novel or emergent properties) participate in, some broader principle of the origination of novelty? In contrast to the law of conservation of energy, which forbids the creation of new energy, creativity provides for the addition of new properties to the universe, over and above anything that was there before.

Gordon Kaufman (2004, p.74) seems to go even further down this road than Whitehead, implying an all-encompassing scope for creativity, covering all forms of cosmic novelty. Kaufman sees human creativity as one of an overall set of three modalities of creativity embracing the entire unfolding of the cosmos[6]:

> Creativity$_1$: 'the initial coming into being of the universe in which we find ourselves, sometimes expressed as the creation of something from nothing (*creatio ex nihilo*).'
>
> Creativity$_2$: 'the creativity manifest in evolutionary processes, the ongoing coming into being of trajectories of increasingly complex novel realities. In this mode creation is not thought of as simply and straightforwardly from nothing (as often with creativity$_1$); it is, rather, creation in the context of other realities that already exist'.
>
> Creativity$_3$: 'human symbolic creativity.'

[6] Here, '*creatio ex nihilo*' is used in a metaphysical context to refer to the divine creation of the cosmos (including the entire material realm) out of nothing. The phrase '*ex nihilo*' is also taken hereinafter in the psychological setting of the individual creative agent, to indicate the arrival, or apparent arrival, of creative novelty or insight out of nowhere, or from nothing, and/or with no clear link to any prior mental content, including pre-existent thoughts or ideas.

These quotes are taken from: *In the Beginning … Creativity* (2004, p.76), where Kaufman also says that 'Creativity$_2$ provides (among other things) a link between creativity$_1$ and creativity$_3$', thereby positioning human creativity relative to evolution.

This highlights the way in which human creativity resembles Darwinian evolution; in both cases the novelty arising through their action was truly absent in any prior state, and it is as if the effect were one of getting something for nothing (the added value from creativity, the adaptiveness of the new variation in evolution). One immediately apparent difference is that in evolution there is no human agent involved; and no designer is required[7]. In this sense artificial selection, which is carried out through the intervention of a human agent, might provide a closer parallel.

Furthermore, just what is the ontological status of creative novelty before anyone has produced it (or even thought of it)? Does it have any meaningful form of existence, or does it not? Were novel properties always existent as latent possibilities, though not present in the environment or matter? Rejecting such pre-existence might be taken as leading in the direction of actualism (denying the reality of alternative possibilities) as concisely summarised by Daniel Dennett:

> This dread hypothesis, the idea that *only* the actual is possible, has been called *actualism* (1995, p.106, citing Ayers 1968).

In an actualist world, creativity seems to become impossible (if not merely irrelevant). Creativity appears to require a range of unactualised possibilities to provide scope for creatively novel output; as Boden has it, creativity needs to produce something '*novel, surprising, and valuable*' (2014, p.227). (Questions of possibility and actuality are discussed further in chapter two).

1.3.5 Creative novelty and inspiration, or intervention

The problems of origination and of the role of the creative agent also have a bearing on the question as to the sources whence creative novelty arises – are psychological

[7] See the discussion in chapter five.

and/or neural processing sufficient, or is there an element of external inspiration, or intervention, necessarily or contingently present? An inspirational source which is external to the creative individual and to the physical universe would simply be beyond explanation in terms of any antecedent causes[8].

Also, in any case of inspirational creativity in this interventionist sense, the inspirational content has somehow to be manifested in an (empirically possible) physical configuration of some kind for it to be brought into actual existence, such that cosmic possibilities are actualised (and perhaps even brought into being) by external agency. This leads to a question of interaction with the physical in the absence of any known channel through which such influence could take effect, much as in the mind/body problem[9] associated with Cartesian dualism (René Descartes proposed a microcosm comprising two distinct substances, mental and physical, which allows for consciousness to be immaterial but in the absence of any adequate means by which the non-physical mind could influence the physical body).

Creativity can and does, of course, disclose new properties hitherto unknown to human culture, and it can be questioned whether the novel output from an instance of creativity is in any sense derived from nothing, or whether recombinations of previously known components are sufficient to explain the new things or ideas, so that the latent possibilities can be actualised, at least in principle, by means wholly internal to the cosmos.

1.4 Towards an account of creative novelty

The previous section has raised a number of questions which any attempt to account for the generation of creative novelty, or to specify a candidate creative process, will need to take into consideration. From the above, the most problematic elements of creativity and the origination of creative novelty appear to be:

[8] Kronfeldner makes a similar point (2009, p.582).
[9] See the entry for 'mind-body problem' in Blackburn (2008, p.235).

1. The opacity of any internal process culminating in the generation of creatively novel thought.

2. The sources of creative novelty and the relation to prior knowledge, belief, or other pre-existing mental content.

3. The more general question of novelty and origination.

4. The relation (if any) between Darwinian theories of evolution and human creativity.

5. The *Meno* paradox and the difficulty of searching for an unknown.

A number of solutions have been put forward at different times and by various theorists, a selection of which are outlined below. Also in this section, I begin to lay the foundations for my own intended approach via Darwinism and mental modularity.

1.4.1 Inspirational, or interventionist, theories

As said, creative novelty can appear to have arisen from nowhere (there may be no clear link to antecedent mentation or knowledge), with any processing involved being opaque to introspection. One possible view is that it does indeed come from nowhere and nothing, or from beyond the physical universe, as in inspirational and interventionist[10] views of creativity. In such cases, the product would not in principle need to have anything to do with previous knowledge, and so intervention by inspiration provides a candidate channel through which novelty might be introduced. Plato took just such an interventionist view:

> … a poet is a light and winged thing, and holy, and never able to compose until he has become inspired, and is beside himself, and reason is no longer in him. So long as he has this in his possession, no man is able to make poetry or to chant in prophecy. Therefore, since their making is not by art, … but is by lot divine – therefore each is able to do well only that to which the Muse has impelled him … . (Plato, *Ion*, quoted in Rothenberg & Hausman, 1976, p.32).

[10] or 'providential', to use Nickles's term (2003, p.55).

Thus Plato held that creativity required inspiration from some source external to the self. Even if this does not amount to an explanation, it is at least in keeping with the apparent impenetrability of creative thinking to introspection. Plato also regarded creativity as incompatible with reason; in this might perhaps be seen an echo of the earlier discussion of creative thinking and novelty relative to problem-solving, algorithmic procedures, and deductive logic.

1.4.2 Creativity and subliminal mental activity

In later times, Neoplatonic and Renaissance views of creation and creativity generally followed the interventionist, Platonic style of account. These theories are non-naturalistic, but some of the threads that are found among them, in particular references to sleep, dreaming, and the suppression of reason, may have relevance even in a more modern context.

Maria Ruvoldt describes the 'sleep of reason' (2004, p.6), the Renaissance notion that reason (as in Plato) acts as a constraint on the reception of inspiration. In sleep (and in dreaming), the effect of this restriction was thought to be reduced, see for example Ruvoldt (2004, p.2). Thus the Renaissance philosopher Marsilio Ficino viewed sleep as promoting creativity (Ruvoldt, 2004, p.22). Mental states facilitating the suppression of reason were also thought to include melancholy and drunkenness (Ruvoldt, 2004).

This long tradition indicating a relation between creativity and states of restricted awareness such as sleep (which is also alluded to in Boden's '*the bed'*), dreaming, and the more general suppression of reason, is sufficiently persistent and well-established to merit some attention, even if in modern times it might not always be interpreted as allowing the apprehension of inspirational material from some external origin.

1.4.3 Theories favouring sources in pre-existing mental content

Unlike Plato, Aristotle regarded the generation of novelty as sourced from the rearrangement of pre-existing constituents:

> 'Now, of things that are being produced, some are produced by Nature, and others by Art, and others from Chance. All things, however, that are produced are produced by means of something, and from something, and become something.' (1991, 1032a, p.141).

Thus Aristotle was more inclined towards the explicability of creativity in terms of the operation of natural laws. The inclusion of 'Chance' as a mechanism for the production of novelty is interesting, given the role of random events in Darwinian evolution, and modern theories of creativity based on Darwinian analogies.

In *The Discarded Image* (1994, subtitled: *An Introduction to Medieval and Renaissance Literature*) C.S. Lewis comments on the transition to modernity, in the course of which the old geocentric models of the cosmos came to be so 'discarded'. Two aspects of the change of model are of particular significance for theories of creativity. Firstly, Lewis identifies 'Internalisation', in which some factors which had once been thought of as extrapersonal influences came to be seen as internal, psychological effects, in particular:

> In this great change something has been won and something lost. I take it to be part and parcel of the same great process of Internalisation[...] which has turned *genius* from an attendant *daemon* into a quality of the mind. (1994, pp.214-215, footnote omitted).

Modernity therefore has much more difficulty accommodating intervention or inspiration from some external source, because, for one thing, in a material universe, there is no obvious channel through which it could take effect.

Secondly, in the medieval cosmology, there was no equivalent of Darwinian-style evolution:

> But the change of Models did not involve Astronomy alone. It involved also, in biology, the change – arguably more important – from a devolutionary to an evolutionary scheme […] (the degree of change can be gauged by the fact that *primitive* is now in most contexts a pejorative term). (1994, p.220).

In terms of the implications for creativity theory, the recognition of evolution by natural selection afforded a greatly enlarged scope and role for novelty in the development of the natural world. Also, in the absence of the guiding hand of a designer (and even

allowing for epigenetic influences) much of this newly recognised novelty arises from random variation and mutation (recalling the Aristotelian ‘Chance’).

1.4.4 The sources of creative novelty

Rothenberg and Hausman note the differing views of Plato and Aristotle, which they regard as being among the ‘seminal accounts’ of creativity (1976, p. xii).

If creative novelty is truly inspired, as in Plato, then there is no natural source, no process is involved, and its origin is not explainable in naturalistic terms. If, alternatively, it derives from something already existing, whether internal or external to the agent, then it might be expected that some process, some series of steps, takes place to transform the source material into creatively novel output. Therefore, if a naturalistic explanation is to be sought, it might be assumed that there will have been a source in some prior state, or some rearrangement of pre-existing elements, such that nothing comes from nothing, or nowhere, or through any extra-natural input. In support of this, creative novelty generally has *something* to do with prior knowledge: Kekulé was familiar with carbon chemistry as it was known at the time, for example, and Plato did not write on quantum mechanics (this is explored further in chapter two). On the other hand, creative novelty cannot be entirely pre-existing; it was not *all* there before, or it would not be original.

Thus, if any novel properties are output, they cannot themselves have had prior existence. So how can this be compatible with any explanation in terms of antecedent causes, when those causes would be expected to anticipate the outcome? One candidate response is that the new properties may be emergent, in much the same way that an old-style mechanical wristwatch indicates the time but the constituent components (cogs, springs, bearings and so on) do not, unless they are configured according to exactly the right structure.

If any input sources are pre-existing, then a previously only possible state has been actualised, so that it is as if the cosmos has an (at least effectively) limitless store of possibilities, in addition to the actual. It will be convenient for present purposes to proceed as if this were the case, for the sake of argument, but without making any claim that it refers to any real state of affairs. Alternatively, if there were multiple worlds corresponding to past and future possible states of affairs, along the lines put forward by David Lewis (1986) or as in the many worlds interpretation of quantum mechanics, then that would also establish a range of possibilities. In essence, Lewis proposed a profusion of possible worlds (in addition to the actual) corresponding to states that did not actually exist, but could possibly do so, so that counterfactual propositions that were not altogether false could be true in at least one possible world. In quantum mechanics, certain equations have alternative solutions corresponding to non-actual states. In the (more usual) Copenhagen interpretation, they are all treated as existing simultaneously prior to observation, at which point only one is seen, but they have alternatively been taken to indicate situations applying in alternative parallel universes, leading to multitudinous worlds in a fashion recalling Lewis's theory.[11] All this might be compared with Alfred North Whitehead's views on possibility and the ultimacy of creativity (see for example 1967, p.179) although Whitehead uses 'creativity' in a very broad sense, close to 'novelty', as above.

This line leads to further questions. What could be the input to a process generating creative novelty? Is it some rearrangement of pre-existing mental content, or does it come from elsewhere, and if so, then where? (External sensory input is one possibility, see chapter two). Also, what is the relation of the creative product to any antecedent states? Where was the novel thought before anyone conceived of it? It must have been possible, so how was the possibility identified and actualised? What is the role in the process of the human creative individual or agent? Further, how (by what process, or series of steps) is creatively novel output formed?

[11] Dawkins (2007, pp.409-410) has a brief description of these alternative interpretations.

1.4.5 Darwinian theories and the proposed Darwinian analogy

One potential response is to seek to draw an analogy with Darwinian evolution, which as said above Kaufman includes within his own broader scope of creativity. As Kronfeldner points out, Darwinian evolution eliminates the need for a designer to explain the emergence of new biological forms in nature[12], though it might perhaps be said that Darwinian evolution explains how novel combinations of pre-existing elements and components can come about but has less to tell us about why certain such recombinations give rise to completely novel properties that were not present in any of those components. It is the new properties, and not the recombined constituents, that apparently arise from nowhere, or nothing. This is the crucial issue for any attempted analogy with human creativity, where the position is less clear, and there remains a similar question as to whether the rearrangement of any pre-existing elements of which a creative product might be composed is itself sufficient to explain the generation of novelty.

Charles Darwin himself described the mechanism of evolution by natural selection in the following way:

> If during the long course of ages and under varying conditions of life, organic beings vary at all in the several parts of their organisation … ; if there be, owing to the high geometrical ratio of increase of each species, a severe struggle for life at some age, season, or year … ; then … , I think it would be a most extraordinary fact if no variation had ever occurred useful to each being's own welfare, in the same manner as so many variations have occurred useful to man. But if variations useful to any organic being do occur, assuredly individuals thus characterised will have the best chance of being preserved in the struggle for life; and from the strong principle of inheritance they will tend to produce offspring similarly characterised. This principle of preservation, I have called, for the sake of brevity, Natural Selection; and it leads to the improvement of each creature in relation to its organic and inorganic conditions of life. (1996, pp.104-105).

As set out here, the process seems to resolve into the effects of two constituent forces:

[12] See Kronfeldner (2007, pp.10-11).

1. Natural variation provides a range of characteristics spread unevenly among a population;

2. The more adaptive variants become more prevalent in succeeding generations due to their greater survival potential.

The combined effects of the interaction of these two pressures lead in due course to the appearance of novel forms of life bearing some of the more successful characteristics.

The two aspects, affording on the one hand variation and innovation and on the other, testing and selection through environmental challenge, appear to align reasonably well with the two basic conditions for human creativity (at a superficial level, at least). On this view, the appearance of new forms out of natural variation corresponds to the novelty criterion, and the survival advantage parallels the need for creative products to be of value, or appropriate, in their context.

In Darwinian evolution, there is no human agent, and no designer is needed. Nonetheless, Darwinism offers a way of understanding a process which generates novelty and hence it can provide a candidate analogy for human creativity (as explored in chapter five), and such an approach has already been proposed by some theorists.

1.4.6 The role of environmental challenge

Evolution by natural selection does not require design input, but there are other factors influencing the variation and selection elements. In particular, individuals bearing new variants are pitched directly against their environment, and it is the challenge presented by the environment that drives the selection (see chapter five). Darwin became acutely aware that the 'struggle for life' could be thoroughly unpleasant for any sentient beings that became directly involved:

> With respect to the theological view of the question. This is always painful to me. I am bewildered. I had no intention to write atheistically. But I own that I

> cannot see as plainly as others do, and I would wish to do, evidence of design and beneficence on all sides of us. There seems to me too much misery in the world. I cannot persuade myself that a beneficent and omnipotent God would have designedly created the Ichneumonidae with the express intention of their feeding within the living bodies of Caterpillars, or that a cat should play with mice. (1860, pp.311-312).

John Stuart Mill, whose essay on *Nature* was written before *The Origin of Species* had been published, also drew attention to the prevalence of suffering within the biosphere:

> In sober truth, nearly all the things which men are hanged or imprisoned for doing to one another, are nature's everyday performances. Killing, the most criminal act recognized by human laws, Nature does once to every being that lives; and in a large proportion of cases, after protracted tortures such as only the greatest monsters whom we read of ever purposely inflicted on their living fellow-creatures. (1874, pp.28-29).

The examples cited by Darwin and Mill result from interaction between the animals involved and their environment (in which I include predation). Evolution by natural selection operates within the scope of the physical laws of nature, the overall effect of which is to favour the continuation of successful variants. Neither the laws concerned, nor the physical process of natural selection, have any bearing on the experience of suffering in themselves. How such physical events come to be felt as suffering (or indeed how any physical event gives rise to, or is accompanied by, sensation of any kind) is, of course, an entirely different question.

But, it is precisely the interaction with the environment which provides the challenge driving the selection. In spite of the fact that it is not infrequently accompanied by suffering, the environmental challenge seems to point to a function within natural selection, in setting up the testing conditions which facilitate the selection process. It is thus effectively a participant in the development of new forms in nature, part of the equivalent of the input from a designer, and a contributor to the emergence of novelty. Physical laws and processes have no ethical or emotional dimension and are simply not equipped to take any account of their effects on the well-being of individual sentient animals. As above, in this context environmental challenge is itself a driver behind the evolution of more complex and more highly developed forms of life, with better means

of responding to those environmental pressures. As there is no designer, its effects are necessarily random and indiscriminate, and with no inbuilt planning or teleology.

Mark Taylor (2001, pp.121-123 & 135) and Kaufman (2004, pp.89-90) who cites Taylor, make similar points about the relationship between information and noise. The principle of the ubiquity of noise applies to the transmission of information through genes, as much as to any other medium, of course. Here, the noise element includes mutations, which can contribute to the evolution of novel forms and characteristics (see for example Dawkins, 1989, pp.16-17). Such intrusive changes are a driver of the generation of novelty. This is again reminiscent of the Aristotelian 'Chance'. Instances of novelty which prove less adaptive in their context are likely to be eliminated by natural selection, through confrontation with environmental challenge.

The role of challenge in creativity is discussed in chapters two and five. Here, it might be noted that, in a more general sense, the context-sensitive value criterion reflects the implied existence of a situation in which there is a need to be fulfilled, or an advantage to be gained, from taking up the novel idea. Put another way, it is as if something resembling an environmental challenge is inviting a response[13].

1.4.7 Searching and the Meno

Darwinian evolution takes existing material as input, and outputs new properties that were not there before. It draws upon coded data accumulated through previous change, and thus has a different relationship to prior information from any that might hold between creativity and relevant previously acquired knowledge (see chapter five). The *Meno* paradox is not applicable to Darwinian evolution because there is no search or recognition, only variation and selection. In creativity, on the other hand, a search procedure might be relevant to the problem of spontaneity, and the relation of novel properties to antecedent states. A search is not a deductive procedure. It may be routine, but it is potentially capable of retrieving results which can be used in new ways.

[13] See Kronfeldner's definition (2011, p.vii) and chapter two.

Thus in a search, spontaneity might be preserved. Searching might therefore allow for the element of surprise noted by Boden. I develop this theme further in chapter five, ultimately leading to the notion of the *requirement* (see chapter six).

1.4.8 Mental architecture

In human creativity, any processing takes place within a human creative agent. In order to investigate the type of process that might be involved, it will be necessary to consider mental architecture, the structure against which that process will operate.

One candidate structure is offered by the modularity thesis (see chapter three). The assumption of a modular framework for mental processing will allow for the construction of a model for a process (in chapter six). The characteristics of mental modules as originally put forward by Jerry Fodor[14] include encapsulation, which would isolate the processing within the module. This might be compatible with the opacity of creative processing to introspection, and with the spontaneous, surprising, and apparently-from-nowhere aspects of much creative output.

Fodor restricts modularity to peripheral input and output subsystems. Other theorists have since proposed extending modularity to more central areas of the mind; this is termed massive modularity (see chapter three). Peter Carruthers is among those advocating such extension; he also writes on creativity, and for this reason his approach is covered in some detail[15]. Carruthers bases his theory of creativity around pretence, rather than Darwinism; this is discussed in chapter three. He applies Darwinian theory in a different setting, evolutionary psychology, in support of his views on massive modularity; his argument is presented in outline, and discussed, in the section below.

[14] In *The Modularity of Mind* (Fodor, 1983).
[15] (In chapters three and six).

1.5 Evolutionary psychology

Darwinian forms of explanation may have further points of relevance to theories of creativity, over and above the provision of a candidate analogue process. In particular, it can be argued that the very existence of human creativity must be due to its conveying some form of evolutionary advantage, either now or at an earlier period during the emergence of the human species. This, evolutionary psychology, view sits particularly easily with the notion that the mind is – at least to an extent – compartmentalised into more-or-less hard-wired modules (see chapter three) as the Darwinian case can then be focused on the development of the various components of an innate mental architecture.

Thus in terms of the current study, evolutionary psychology might be helpful in providing a possible argument for modularity (which is to be taken as a framework for the modelling exercise) and by extension to explain the development and function of human creativity, to which it is particularly relevant (as opposed to artificial creativity, or to an understanding of creative process regardless of platform).

The evolutionary psychology approach is in opposition to the standard social science model, according to which the mind is initially *tabula rasa*, without any fixed set of functions determined from birth, but with mental development occurring through exposure to the cultural environment (see for example Mithen, 1996, p.9). As a result, evolutionary psychology is controversial and far from universally accepted, and so this line must be pursued with caution.

Evolutionary psychology is a very much an active field of enquiry, with an extensive literature, but for current purposes, it need only be taken into account in so far as it might throw some light on the creative process. Thus at this point discussion will be restricted to consideration of a single author, Peter Carruthers, and more specifically to the argument set out in his 2002 paper, which incorporates a concisely expressed example of an evolutionary psychology approach. It should be borne in mind that

Carruthers has published further works in this area since then – see, for example, *The Architecture of the Mind* (2006), in which he updated his opinion – although here I shall focus specifically upon the argument from evolutionary psychology as it appears in the earlier paper. Carruthers does not draw a direct analogy between creativity and evolution, but his assessment of the cognitive requirements for creativity is important for his understanding of the creative process. (Carruthers' views on creativity will be explored further in chapter three).

Carruthers states in the abstract accompanying the paper that he is setting out to defend the following claims:

1. 'that essentially the same cognitive resources are shared by adult creative thinking and problem-solving, on the one hand, and by childhood pretend play, on the other – namely, capacities to generate and to reason with suppositions (or imagined possibilities).' (2002, p.225).
2. 'that the evolutionary function of childhood pretence is to practice and enhance adult forms of creativity.' (2002, p.225).

He then proceeds to take these claims as the foundation to support a proposed explanation of the interval between the emergence of the modern human species in Africa about 100,000 years ago, and an apparent global 'creative explosion' (but see below) of rapid development which was not thought to have occurred for another 60,000 years.

Carruthers bases his explanation of the evolutionary roots of creativity on his proposed link with childhood pretend play, which is common across all cultures (2002, p.228), although the young of other mammals do not show this tendency. Carruthers suggests that creativity and pretend play both have a basis in the imagination (and thus involve exploring possibilities beyond the actual situation). He proposes that childhood pretence practises for creativity in the adult (in the way that kitten play is practice for feline adult skills such as hunting). He argues that both of these activities require: 'a capacity to generate, and to reason with, novel suppositions or *imaginary scenarios*' (2002, p.229).

In assessing the 'cognitive requirements' for pretence, Carruthers cites Shaun Nichols and Stephen Stich's (2000) model of pretence[16], which he says would call for an architecture that incorporated:

1. '...a supposition-generator, or "supposer"[17], to create representations of imagined possibilities...' (2002, p.241);

2. '...a "possible worlds box", which is a memory system designed to store suppositions and their elaborations during the course of a pretend-episode.' (2002, p.241)

Carruthers considers that established memory arrangements would suffice to house the possible worlds box, and he suggests that initial suppositions could be supplied by two subdivisions of imagination:

1. '*experiential*' operating with images (not restricted to visual) (2002, p.241)

2. '*propositional*...the capacity to form and consider a propositional representation without commitment to its truth or desirability.' (2002, pp.240-241)

He argues that both would have been available 100,000 years ago, along with fully human language.

The formulation of mental content by these two faculties would not, as Carruthers says, by itself be adequate for what he terms 'imaginative thinking', which he sees as 'consist[ing] in the generation of *relevant* and/or *fruitful and interesting* new ideas.' (2002, p.242) It might also be noted that this would then meet both the novelty and the value criteria for creativity, as introduced earlier. Hence Carruthers proposes that the function of pretend play is to 'help build a consistent disposition to generate novel suppositions which *will* be both relevant and interesting.' (2002, p.242).

[16] Nichols and Stich incorporated a revision of their 2000 paper in Nichols and Stich (2003), which is the version subsequently referred to here.
[17] Nichols and Stich (see 2003, p.35) refer to 'the *Script Elaborator*'. Both Carruthers (p.241, footnote 11) and Nichols and Stich (2003, p.35, footnote 8) acknowledge that the script elaborator may initiate a pretence as well as fleshing it out.

The timeline Carruthers takes as a reference for comparing the explanatory power of various theories with respect to the fossil record discerns evidence of 'modern intelligence' from as early as 90,000 years ago, although at that stage without those signs of creativity which appeared later on[18]. The tool-set in use continued unchanged until the commencement of the 'creative explosion' about 40,000 years ago, when new technological and cultural developments became reflected in fossil remains[19]. By then, human populations were dispersed around the globe, and yet it apparently took only a further 10,000-15,000 years for these advances to become worldwide. Carruthers argues that it is unlikely that this reflects a global genetic change, because in that case, differential development would be expected in dispersed and isolated communities, which is not seen.

In sum, Carruthers' view is that pretend play and creativity were both feasible from 100,000 years ago. 'But it took 50,000-60,000 years for a strong disposition towards pretend play in childhood to become established in the human phenotype...' (2002, p.242) that is, culturally, without accompanying genetic change being necessary. This development was driven by the pressure for the enjoyment of pretend play, together with selection, sexual or environmental, favouring the disposition to indulge therein (because of the creative intelligence engendered in the adult).

As said above, Carruthers has since revised his position. In the 2002 paper, in particular, it was accepted as a constraint on any theory of creativity that it should support an explanation of the aforesaid 'creative explosion'. But by 2006 (in *The Architecture of the Mind*, see pp.278-280) he had become inclined to dismiss the creative explosion as an artefact of: the European focus of archaeological research; the low European population (there having been no modern humans in Europe until about

[18] Carruthers cites Mithen (1996) as source (p.227).
[19] Carruthers cites Stringer and Gamble (1993) and Mithen (1996) as sources (p.227).

40,000 years ago); and also the gaps in the fossil record (for example, creative thinking when tracking prey during hunting[20] would be undetectable).

Nonetheless the line of argument in the paper successfully relates the development of creativity to an evolutionary context, and for current purposes demonstrates, at the very least, how it might be achievable to dovetail creativity theory with Darwinian logic.

1.6 Language and creative thought

Carruthers (2002) rejects the notion that language alone would have been sufficient as a foundation for the emergence of creativity. In developing his argument, he examines a few alternative accounts of human creativity, together with their implications for its evolutionary and cognitive bases. One of these, from William Noble and Iain Davidson (1996), has fully syntactic natural language as sufficient for creativity. On Carruthers' timescale, this would imply that full language did not appear until 40,000 years ago (although he thinks that unlikely due to the then widespread geographical dispersion). Carruthers notes that Noble and Davidson suggest languages are an expression of a more general "symbolic capacity", but finds that this runs counter to Chomsky's well-supported theory of innate language capability, and that the global dispersion would argue against its being spread through social interaction. Carruthers thus concludes this view is 'implausible' (2002, p.232).

Instead, he contends (as said above) that the propositional and experiential powers of imagination would together have been sufficient for the generation of suppositions on the Nichols and Stich (2000)[21] model. He argues that a supposition generator 'would have required the creation of a whole new type of propositional attitude' (2002, p.241), if it were not to be available in this way via the two sub-faculties of imagination. Presumably the propositional attitude in question would be that of supposing, as might be represented by *to suppose that P*. And as Carruthers says:

[20] See Carruthers (2006, chapter 6).
[21] Nichols and Stich incorporated a revision of their 2000 paper in Nichols and Stich (2003): my footnote earlier in this chapter refers.

> … pretending is a kind of *supposing*. To pretend that P is to suppose that P for fun, roughly speaking. And sustained episodes of pretence are best understood as episodes of thinking and reasoning within the scope of a supposition.[…] (2002, p. 236, footnote omitted).

Even though Carruthers rejects its emergence from scratch, it might be recognised at this point that this propositional attitude presents a formidable tool for creative mentation, in that: *to suppose that P* could seed a mental model of a possible world in which P is the case (much as *to believe that P* constructs an element of a model of the actual world), although it is difficult to see how either Carruthers' 'capacity to frame and then consider a possibility' (2002, p.242), or the ability *to suppose that P*, might be sufficient to explain where any novel ingredients of 'P' could be sourced, any more than the existence of Carruthers' twin faculties of imagination could by itself explain how the content of appropriate suppositions is to be procured (even given that they are able to hold and manipulate the types of data necessary to build suppositions, to elaborate them, and to construct models of possible worlds).

Whether or not on its own it is sufficient for creative thinking (and whether or not it formed the original evolutionary vehicle for creativity), language clearly does allow for the assembly of novel sentences and for reference to non-actual and counterfactual states of affairs. Vocabulary and rules of syntax provide for the generation and exploration of such possibilities, supplying a basis for the construction of suppositions (in Carruthers' terms), if not in themselves promoting appropriateness of content.

Also among the facilities that language allows is the transmission and reception of information 'about the not-here-and-now' (2002, p.235), and Carruthers notes the significance of making mental models in the interpretation of such communication. He cites Paul L. Harris (2000) as arguing that pretend play exists to practise the construction of mental models, but he also observes (2002, p. 236) that pretence requires a model of a situation that is not only not-here-and-now, but beyond the actual world altogether. Carruthers notes that pretence is thus in effect more highly powered than is strictly needed for the task of understanding testimony alone (and so its

development is not fully explained by the need for such comprehension), although it is well suited to practise for creative thinking, which clearly does require the imagination of non-actual possibilities.

Nevertheless, this might be an appropriate juncture to introduce the potential role of language in creativity, given that it is capable of framing statements about non-actual, but possible, states of affairs (not to mention impossibilities and nonsense) – the recognition of a novel possibility, previously unknown to the originating individual, being a candidate requirement for creativity.

Nor are these facilities all that language has to offer as a potential medium for creative processing, for besides the meaningful content of statements (whether or not representing some aspect of the actual world), language brings with it the properties of the symbols by means of which that content is expressed. From the symbols, scope for further relationships opens up between words and phrases. These include homonyms, synonyms, and antonyms of course. But, particularly when written language is available, further features of words come into play. such as number of letters, similarity of letters and/or sounds, rhyming, and so on, which are unconnected with the meanings that they are intended to convey or the things to which they are taken to refer. David Perkins writes about these additional properties and relationships in *The Mind's Best Work*, and gives as an example:

> White, soft, and edible are properties of spaghetti, … and certainly not properties of the word "spaghetti." In contrast, "starts with the letter s" is a property not of spaghetti but of the word that names it. (1981, p.76)

The additional relationships greatly increase the capacity for linkages between words and phrases, whether or not connected by meaning, thus adding to the potential offered by language as a vehicle for creativity, and perhaps also as supplying some of the facilities that might underpin a mechanism for a creative process.

1.6.1 Metaphor and analogy

Further opportunities are provided for the linking of disparate words, phrases, and meanings by the use of figurative language and expressions. Metaphor and analogy allow for the connection or association of ideas from two or more (sometimes widely separated) spheres of thought. Perkins (1981, p.85) notes the widespread observation that metaphorical thought is frequently entailed in discovery and invention. He gives the example of Kekulé's work on the ring structure of the benzene molecule (see below). However, Perkins says that he scrutinised a number of reported cases of insight, but concluded that: 'The surprising result was that novel, remote analogy seldom – not never, but definitely seldom – contributed to these insights.' (1981, p.89).

Perkins also discusses (1981, pp.91-99) Arthur Koestler's theory that creativity involves the 'bisociation' of different matrices of thought, from *The Act of Creation* (1964). But Perkins notes that 'Bisociation seems … to be defined largely in terms of its products' (1981, p.93), and he concludes that it does not necessarily correspond to any one process by which those products are reached.

The possible relevance of these areas of language, metaphor, analogy, and association will be explored in more depth later; for current, introductory purposes it is sufficient to note their presence, or existence and potential, as areas for consideration. The question as to what kind of process or method might be expected to lead to a creative outcome will also be pursued in subsequent chapters, but for the moment mention needs to be made of one frequently adopted model, due to Graham Wallas in *The Art of Thought* (1926).

1.7 The four stage process model

Wallas's model has been taken up by Catharine Patrick (1955) and Jacques Hadamard (1949), among others, and comprises four stages[22]. Wallas (1926, pp.80-86) names the stages as:

1. Preparation
2. Incubation
3. Illumination
4. Verification

Preparation covers all the groundwork in acquiring the knowledge and skills necessary for the insight in question. For example, Kekulé came up with his structure for benzene only after much previous work on the make-up of organic molecules (see Findlay (1965)). Similarly, neither Plato nor Aristotle put forward a theory of relativity, even though they manifestly had great intellectual capability: they would not have been able to do so without access to the enormous amount of prior work and understanding in physics previously assembled by Isaac Newton among many others. Newton (1676) himself said, 'If I have seen further it is by standing on the shoulders of giants', and it is necessary first to climb upon those shoulders in order to reach higher still.

The second stage often corresponds to a period when no active work is being undertaken on the issue in question, and perhaps little progress is apparent. Wallas refers to Helmholtz: 'the second is the stage during which he was not consciously thinking about the problem, which I shall call Incubation' (Wallas, 1926, p.80). But this step is the precursor to the third stage of illumination.

[22] The first three Wallas derives from Helmholtz (via Rignano, 1923, pp.267-268; Rignano cites Helmholtz, 1891). Wallas (1926, pp. 75 & 80) also refers to Poincaré's *Science and Method* (the relevant section appears in the Halsted translation under *Mathematical Creation*, pp.383-394, see Poincare (1908)), and discerns the four stages in Poincaré's recollections. See also chapter five.

Illumination is the point at which the new, critical insight is achieved. This is frequently reported to have happened when the individual concerned is not actively engaged with the problem. Boden's reference to '*The bath, the bed, and the bus*' (see above) might be recalled in this context. This stage might also be taken to tie in with the 'novelty' aspect of the criteria for creativity.

The final stage Wallas calls 'verification', which Patrick expands to 'verification or revision' (1955, p.41). The initial insight may not provide a complete solution to the problem. It may be necessary to reverse engineer the new idea to some extent to fill in the logical steps leading up to the insight, of which the creative individual may not be fully aware at the time of the illumination. It may require some work of elaboration to get the result into a state in which it is useful in a practical application context. It is also necessary to check the validity of the insight, that it holds good under test, that it is capable of meeting the criteria sought. This stage may therefore be seen to relate to the 'value' aspect of a creative product.

Paisley Livingston mentions the four stages in *Art and Intention*. (2005 pp.47-48 and footnote 33). He seems a little doubtful about Wallas's sourcing (of the first three) from Helmholtz, although Wallas quotes from the (admittedly very brief) passage as it appears in Rignano. Livingston also notes possible alternative derivations of the stages, other than via Wallas. Kronfeldner (2011, pp.39-40) gleans four very similar stages direct from Poincaré (see chapter five). But whatever their provenance, the stages seem in any case to be anecdotally, rather than evidentially, based. However, they have gained wide acceptance.

Much more will be said about creative process in later chapters, but here it is sufficient to see that, even without going into detail on mental structure and functioning, it has been possible to suggest a few elements which might be present in at least some instances of creative thought and action.

1.8 Summary and conclusion

This brief preliminary excursion into the characteristics of creativity and some of the problems, issues, and questions associated with them has yielded a few particularly significant points, which can now be summarised.

Creativity in the sense taken here is a widespread human activity, which is chiefly distinguished by the generation of novelty and added value; both are necessary criteria, but they may not be sufficient on their own, as will be explored further in the next chapter. As a channel for novelty, creativity can have ontological significance.

Human creativity involves an agent, a process, a product, and also (conjecturally) source materials that do not, in themselves, carry the novel aspect of the creative output. The novelty condition requires that a creative product cannot be known in advance, and the operation of the creative process tends to be obscure.

It is assumed that creative products are the effect of antecedent causes that are (in principle) knowable, and that a creative process does not require intervention; these assumptions are conditions that must be satisfied for explanation to be feasible.

There is a widely accepted sequence of stages which are regarded as generally applicable to a creative process.

In that novelty is generated, creativity has some claim to comparability with Darwinian evolution by natural selection, the direction of which is driven by environmental challenge.

The other link with Darwinism is that there have been attempts to explain the development of human creativity via evolutionary psychology. Carruthers, in particular, brings out the commonality between creativity, supposition, and pretence. There is also a question as to how closely creative process is linked to language, and/or as to how language development facilitated the development of human creativity.

These points will be taken forward as and when appropriate in subsequent chapters. They indicate specific questions and features of creativity that will need to be catered for in chapters five and six, which investigate the feasibility and construction of a model for a creative process.

In the course of the above discussion, a few examples have been mentioned of introspective reports by creative individuals of instances of creative insight. Space and time will not allow any extensive descriptions, but it may be helpful to present one or two here in outline form. Accordingly, the following section appends a few summaries, together with references indicating where further such accounts may be found.

1.9 Examples of introspective reports of creative insight and serendipitous discovery

Purely mental events are, of course, not publicly observable, and so introspective accounts are by their nature unverifiable. There is no knowing how or whether the content of what is recalled actually relates (or does not relate) to any mental process central to creative thought. This does not render them worthless, because they may provide pointers to elements in creative process that might be replicable in other ways. In the absence of being able to observe creative process at a neural level, these reports may be the closest approach attainable to observation of creative process in action, for all their drawbacks.

1. Friedrich August von Kekulé

The first example relates to Kekulé's discovery of the molecular structure of benzene. His account is very widely cited in the literature, and also at various points in this and subsequent chapters. There is a good account in Findlay (1965, pp.39-40). Kekulé describes a daydream in which chains of atoms came to resemble snakes, one of which grasped its own tail, thus forming a closed loop (benzene has six carbon atoms arranged in the form of a ring).

2. Henri Poincaré

In *Science and Method*[23], Poincaré relates that, having drunk some coffee, he had an episode of insomnia, during which he experienced numerous ideas appearing and combining with each other. In the morning, he was able to establish a set of mathematical formulae.

3. Samuel Taylor Coleridge

Coleridge (who was unwell at the time) was staying at an Exmoor farmhouse. He reports taking opium and then falling asleep, having been reading about Kubla Khan beforehand. He experienced images and assembled two or three hundred lines of poetry while in that state. On awakening, he set about writing them down, but was interrupted by a visitor from Porlock. He subsequently found that much of his recollection of the new material had faded. (See Jackson, 1985, p.102 and Ghiselin 1985, pp.83-84). This summarises *A Fragment*, prefatory to *Kubla Khan*, but it should be borne in mind that Jackson, in his notes to the text, suggests that Coleridge intended the *Fragment* to be taken as an indication of 'the frame of mind in which this poem (which Coleridge called a "psychological curiosity") was to be read', rather than as representing any plain factual account. (1985, p.702).

4. Alexander Fleming

Fleming's discovery of penicillin is an example of serendipity (see Roberts 1989, pp.159-164). Serendipitous discovery involves the observation of chance environmental clues, which are used to put together creative solutions to (not necessarily connected) problems. In brief, Fleming, who was working in a laboratory at the time, noticed that a petri dish containing a culture of bacteria had become contaminated by mould, which had resulted in part of the dish being free of the

[23] See Poincaré (1908), also Ghiselin (1985, pp.22-31), and Wallas (1926, pp.80-81).

bacteria. Fleming then proceeded to investigate the bactericidal properties displayed by the mould, leading to the identification of penicillin.

Further reference

Additional examples, together with much related material, may be found in: Burke (1978), Ghiselin (1985), Roberts (1989), and Vernon (1970).

Chapter Two: *Creativity and Context*

2.1 Introduction

In the previous chapter, I explored the general field of creativity, and some of the questions associated with it. As a result, I am now in a position to focus more directly upon the essential characteristics of creativity itself, and the kind of process that might be capable of mediating it.

I will begin by looking more closely at the range of definitions and descriptions that have been put forward, and the varieties of creativity (and of novelty) which some theorists have identified. I will also discuss its relation to nature and human culture, and to possible and actual modes of being. In this, I will not so much be attempting to construct any precise, comprehensive, or definitive statement of its meaning, as to form as rounded a picture as possible of its attributes, in order to establish a sound base for modelling by building a thorough understanding of what it is that needs to be modelled. (In any case, creativity might not be a unitary concept, nor is it necessarily always mediated via a single process.)

Towards the same end, I will consider problems of explanation of the origin of novelty (novelty invariably being generated in any creative act) and its relation to any prior state. I will revisit the *Meno* paradox and the problems of searching for an unknown and of recognising it when found. I will discuss the relation between this and the process of evolution by natural selection. These investigations will reveal aspects of creativity with specific implications for modelling, and for the treatment of the proposed analogy with natural selection.

The current chapter therefore begins with my seeking a more complete and detailed understanding of the nature of creativity. I will need to narrow the field down to something sufficiently plain and straightforward for modelling, while ensuring that the target is not too far removed from what is generally understood to be creative output.

I bring in the agent, process and output components pertaining to an instance of creativity, the output being the obvious target for definitions because it must be empirically possible and thus potentially publicly observable. I discuss how it is that novelty and value have been described as 'criterial attributes'. I consider their context-sensitivity, various refinements that might be applied to the two basic criteria, and also Kronfeldner's originality and spontaneity conditions and her reference to environmental challenge. I note that novelty and value are necessary but insufficient for creativity, but argue that they can reasonably be taken as diagnostic for the purposes of modelling.

Novelty I will consider in its ontological context within nature and human culture. Novelty arises universally, so I will need to distinguish creative novelty from all other occurrences. To do this, I will accept the division into novelty arising in nature and novelty arising in human culture (including artefacts), so that creativity originates novelty in culture. I observe that human culture is composed of novel output, not all of which is creative (it may be copied or reproduced from an original).

I discuss relevant aspects of Mill's view of Nature and Art and I note that his inclusive version of Nature takes in possible as well as actual states of affairs, and that the actual world is a subset of all that is possible. I propose that creativity can alternatively be described as making actual something that *was* only possible, within human culture. I discuss the nested series of the degrees of possibility and the requirement for the output of a creative process to be empirically possible (even if it may symbolise higher degrees of possibility).

I will then move on to consider - and to distinguish between - a number of varieties of creativity that have been identified, along with their corresponding varieties of novelty

(historical, psychological, and anthropological). I comment on the differences with respect to macrocosmic and microcosmic viewpoints. I agree with Boden and Kronfeldner that the psychological variety is of primary importance for the understanding of creativity. I note that Kronfeldner adds a metaphysical variety, not part of the nested series formed by the other three; it involves *ex nihilo* generation of novelty and is not amenable to explanation or modelling. Metaphysical creative output would not have complete causal links to any prior state, and I argue that demonstration of a viable creative process would show that non-metaphysical creativity was plausible (though it would not disprove the metaphysical variety).

I also look at Boden's varieties of creativity related to surprisingness (combinational, exploratory, and transformational). I note that they refer to, or depend upon, conceptual spaces which are defined by rules and resemble possibility spaces. I equate the set of all possibility space with my conjectured 'universe of possibility' (for which I make no claim to ontological status). I consider whether transformational creativity may represent a separate, distinct mechanism in that it depends upon prior knowledge of the conceptual space rules, as opposed to prior understanding of the subject matter. I will suggest that Kekulé can be taken as an example of recombination or association, as well as an example of Boden's transformational creativity.

I argue that creativity must depend on previous learning, though not in a direct sense, and that it may be that resultant novel (emergent) properties are not wholly reducible to elements present in any prior state. I identify three possible sources of the input materials to a creative process: internal to the creative individual, in the external world, or elsewhere (in other words, nowhere, or *ex nihilo*). I will set up the assumption that it is as if all possible novelty is already present in the conjectured universe of possibility.

I consider the sources of novelty in evolution by natural selection as being known to be pre-existing physical entities (which are reassembled into novel arrangements in the evolutionary process). I will argue that the input sources for creative novelty must

similarly lie within the actual world, if there is to be no *ex nihilo* element in the process, and if it is to permit of explanation in terms of antecedent causes.

I will then proceed to examine the relationship between the possible (as in the state of potential creative novelty prior to the creative act) and the actual (as in the potentially publicly observable output of a creative process). In this connection, I consider the information capacities of the universe and of the human individual, and I attempt to assess the effect of the extent of the universe of possibility on the scope of potential recombinations of existing elements. This will have particular implications for any candidate creative process which might include a random element, and I take the *Universal Library* as an illustrative example. This leads on to a discussion of the relation between all possible states, no possible states, and the actual world (as the accumulation of novelty). I note that random recombination is undoubtedly capable of producing creative novelty, but that such an outcome is vanishingly improbable in the absence of a creative process. Similarly, the random generation of text produces novelty but the novel text is unlikely to add value to the cultural environment. I observe that all this is a problem for theories of creativity involving randomising (in the way that the generation of evolutionary novelty might call for random mutation).

I note that it is possible to get more out of a random process than was put in (as in evolution) because emergent properties, not present in pre-existing states, may arise when input materials are rearranged into novel combinations. I argue that such emergent properties effectively arise *ex nihilo*, because they were absent from prior states and pre-existing components. I will suggest some equivalence between random mutation (evolution) and serendipitous discovery (creativity).

I will then turn to consider the *Meno* paradox (introduced in chapter one) and its position with regard to evolution by natural selection as a candidate analogue for creative process. I introduce the notion of a *requirement*, based upon prior knowledge of desirable but unavailable properties, as the key to the search and recognition issues

raised by the *Meno* paradox, and which I argue corresponds to a stage that is missing from Plato's account.

I conclude with a summary of possible steps for a creative process: requirement, search and matching, and recognition, which are absent from evolution by natural selection, and would thus require modifications to any strict analogy if it is to be taken as a basis for constructing a model creative process.

The objectives for this chapter

This chapter accordingly focuses directly on the nature and characteristics of creativity itself, and its role in the generation of cultural novelty, which is the core issue for the thesis. The intention here is to identify, isolate, and clarify the key aspects of creativity and the areas with which it overlaps, in order to distil and extract those elements which it will be necessary to explain and/or model in order to understand more of the origination of cultural novelty and any creative process through which it might be mediated.

Accordingly, a selection of different descriptions and definitions of creativity are detailed and assessed. The varieties of creativity and novelty are categorised and related. The setting of creativity relative to nature and culture is discussed. Further consideration is given to the problem of explanation raised by the origination of novelty: if something is completely new, how can it be traced back to antecedent conditions, and to what extent can it be considered to be determined by them? This also relates to the question of sourcing for the constituents of novel ideas. If the creative product is taken to represent a recombination of pre-existing elements, there is then the problem of identifying potentially fruitful combinations from among the (often prohibitively vast) number of possible configurations. The *Meno* paradox, that relating to the problem of searching for something when it is not known either what is sought or where to look for it, and then of recognising it when it is found, is also associated with the origination problem and is aired in this chapter. It is here related to evolution by natural selection,

which will later be taken as a modified analogy for creative process, forming one of the planks on which the modelling exercise will be founded.

Reference will be made from time to time, in this and subsequent chapters, to introspective and anecdotal accounts of creative experience, such as those described in chapter one, although it must always be borne in mind that unobscured and unambiguous, reliable and verifiable, direct insight into the actual mechanism by which a new idea is brought forth is simply not achievable by such means.

These investigations reveal various aspects of creativity which may have implications for any attempt to construct a model for a creative process, given the potential need to be compatible with, or to replicate, any or all of them.

2.2 What is ‘creativity’?

Chapter two thus starts from the initial description of creativity established earlier, and expands on this to examine its qualities and attributes from a number of different perspectives.

The nature and scope of creativity were discussed in brief at the beginning of chapter one; this section begins with a more detailed exploration. It will be essential for what follows in the modelling exercise to understand as fully as possible what is meant by ‘creativity’, so that reasonable and appropriate expectations can be established for the modelling exercise. It will be helpful to distinguish, as far as practicable, the core attributes that must be reproduced and those that may be more peripheral, and perhaps effectively optional. This will be an important step in the direction towards ascertaining or appreciating what exactly will need to be modelled, what it is that requires explanation, and whether any features can safely be ignored for such purposes.

A creative act was earlier taken to involve four components: the (conjectured) source material, the agent (the creative individual), the process, and the product. Definitions

and criteria could be focused on any or all of these. The source materials are less clearly significant than the others for this purpose as they are purely input to the process, and hence take no further part in the action. The output, which is the seat of the basic qualities of novelty and value, is the most obvious target, due to being publicly observable, as previously said. The process and agent are a degree more problematic. The process is largely unknown. The relationship between agent and process is unclear: is the process in any sense within the agent, and to what extent does the agent initiate, control, or otherwise influence the process?

From the first chapter, novelty and value were seen to be (effectively) necessary attributes of creativity, although their evident insufficiency hinted at further complexity and fuzziness around the edges of the concept. There is therefore a need to narrow the frame to something sufficiently well-defined for modelling and then to check that what is taken as the target is not too far removed from creativity or a creative process as they are normally understood.

There follows a selection of descriptions and definitions by various theorists, to illustrate the range.

2.2.1 Novelty and value and their relation to context

To recap briefly, to ‘create’ something is to bring it into existence. ‘Creativity’ carries a similar implication of the introduction of newness, although here tending more towards the sense of ‘originality’. Further, novelty alone is inadequate, however original; ‘creative’ only applies to *some* instances of the generation of novelty. This is mainly due to an underlying presumption of added value, as in Rothenberg and Hausman, who take it that ‘newness and value’ are ‘essential’ and ‘criterial attributes’ (1976, pp.6-7).

Novelty is clearly relative to that which has previously come into being: if it already exists, it cannot be new. It is also partly conditional upon the possibility of awareness

of any prior or parallel instances of the novel idea or entity; such developments may have occurred beyond the agent's knowledge horizon, as Dean Keith Simonton notes:

> What may be original with respect to one culture may be old news to the members of some other culture. Thus, Galileo's discovery of sunspots counts as an original contribution to European civilization even though the Chinese had noted their existence for well over a thousand years. (1999, p.6).

This question is discussed further in terms of the historical and psychological varieties of novelty, below.

The value criterion is also dependent on context. Just as new biological variants have to contend with the current range of environmental conditions and challenges, so do creative products have to add value against the background of the cultural and artificial environments (as well as the natural). Cultural and natural environmental conditions vary with place and over time. Just as the rat, which is successful in the current environment, would have failed in primitive earth conditions with a reducing atmosphere, creative products are not necessarily appropriate to all times and places. For example, a van Gogh painting is much more widely appreciated (and more highly sought after) now than it would have been during the artist's lifetime; tastes and fashions change.

> Even the most shocking novelties will be somehow related to an old style. This relationship may not be immediately apparent, and still less well understood. Acceptance of the new structures as being "valuable" may take some time, for the stylistic similarities and differences must be both recognized and tolerated. (Pablo Picasso's cubist canvas *Les Demoiselles d'Avignon* was initially rejected even by his artist friends; when it was finally exhibited a few years later, it caused a public scandal.) (Boden, 2014, p.229).

Also, potential creative products may fail due to extrinsic factors, such as the agent's deciding not to express or implement the idea, or due to expression being prevented by political pressure or accident (see the note on Coleridge at the end of chapter one, for example). There may therefore be a case for amending the value criterion from 'adds value' to 'propensity to add value', following the pattern that has been suggested for 'fitness' in biological evolution (see the section on the tautology problem in chapter

five). But this would only be a partial fix, because propensity to add value can only be assessed in the light of what is known about past and present cultural and natural environments. Unforeseen developments may render a currently unrecognised idea of value at some future time, and hence lead to its subsequent reclassification as 'creative'.

Thus, there is no absolute standard against which to measure the degree of added value attaching to a candidate creative product, in the same way that there is no standard set of environmental conditions to test the success or otherwise of a new genetic variant.

For these and other reasons, novelty and value can each benefit from a degree of refinement. There is a variety of opinion as to the appropriate qualifiers, as will be seen from the examples below.

Kant

Novelty (as originality) and value are generally seen as necessary criteria for creativity, but even taken together, they are not always regarded as sufficient in themselves. Kant, for example, says that creative output has to be '*exemplary*' that is, it needs to have a touch of the seminal, or of the establishment of a paradigm (1793, Ak.307-308, p.175).

Boden

Boden requires the novelty to have an element of the unexpected or unforeseen:

> My own definition sees creativity as the ability to generate creative ideas (a shorthand term that includes artifacts) – where a creative idea is one that is *novel, surprising, and valuable*. (Boden, 2014, p.227).

Sternberg and Lubart

Robert J. Sternberg and Todd I. Lubart highlight the context-sensitive aspect of the value criterion, by recasting it in terms of appropriateness:

> Creativity is the ability to produce work that is both novel (i.e., original, unexpected) and appropriate (i.e., useful, adaptive concerning task constraints)... (1999, p.3).

(The adjective 'adaptive' is here reminiscent of natural selection – Darwinian theories of creativity are discussed later in this chapter).

Novelty and value both refer to the output of creative activity. Also required are: the human input (the agent), and the means by which the product is brought forth (the creative process, the procedure by which the output is generated), as previously said. Some theorists include properties of these in their definitions of, or criteria for, creativity. (The question of the raw material out of which the creative product is formed will be explored in detail later.)

Gaut

Berys Gaut brings in the agent and the creative process. He excludes routine or fortuitous production, and calls for the manifest exercise of creative talent:

> … And we must, at least, rule out cases of making by chance or by mechanical procedure, if an act is to count as creative. I will say that the making must involve *flair* by the maker to rule out at least these kinds of cases. (2003, p.151).

Carruthers

Carruthers, in his 2006 book *The Architecture of the Mind*, begins the chapter on creativity (Chapter 5: *Creative Cognition in a Modular Mind* (2006, pp.277-334) by considering its features, qualities, and attributes. He includes a reference to the actions of the creative individual:

> Creativity, as I shall understand it, is constituted (at least in part) by a capacity to combine ideas in novel ways in abstraction from any immediate environmental stimulation. And it will normally manifest itself in new types of behaviour, going beyond mere reapplications of

> established scripts or action-patterns. So anyone who is imagining how things could be other than they are (and other than they can be predicted to be) will be thinking creatively. (2006, p.277)

This binds the output closely with the agent. 'Thinking creatively' here seems to be distinguished from 'creativity', which can encompass putting new ideas into effect.

Carruthers' description of creativity preserves the requirement for newness, but does not incorporate explicit reference to added value. He had previously used very similar, partly behavioural, terms in his 2002 article (p.226), but later in that paper he introduces a further concept that he calls 'imaginative thinking', which brings in the value condition.

> Of course a mere capacity for creative generation of new sentences (or images) will not be sufficient for imaginative thinking as we normally understand it. […] Rather, we think imagination consists in the generation of *relevant* and/or *fruitful and interesting* new ideas. (2002, p.242).

Kronfeldner

Kronfeldner, near the beginning of the 2009 paper *Creativity Naturalized*, observes the widespread use of the two basic criteria for creative output, following Sternberg and Lubart in taking these as novelty and appropriateness. She registers the need for further elucidation of these twin attributes, also citing Boden's "…*new*, *surprising* and *valuable*." (2009, p.578, and Boden, 2004, p.1). Kronfeldner proceeds to focus on the novelty aspect, and argues that it has to have qualities of originality and spontaneity, as discussed in detail later.

In *Darwinian Creativity and Memetics* (2011), Kronfeldner again acknowledges the widely accepted characteristics of novelty and value, and also relates them explicitly to the external environment:

> Creativity, in its basic sense, is the human capacity to create new and valuable responses to challenges to which humans are exposed, or to which they expose themselves. (2011, p.vii).

This definition refers to creativity as a general human tendency, rather than focusing on particular instances of creative thought, but retains a reference to the two basic attributes, as said. The link to environmental challenge is however a thought-provoking addition; creativity will be related to responsiveness to external factors in a not dissimilar way in the modelling exercise, later on. It also recalls the discussion on the role of environmental challenge, in chapter one. The cultural, artificial, and natural environment, with its challenges and niches, is the context against which the value added by creative novelty is judged, as discussed above.

Novelty, value, and other indicators

Creativity involves agent, process, product and perhaps source. Descriptions and definitions of creativity were most commonly seen to refer to the product, and specifically to the novelty and value criteria as applied to the product. It appears from the above that both novelty and value are generally taken to be required for creativity, although it also seems clear that even both together are not usually regarded as sufficient indicators. Nonetheless, after this consideration of a number of authors, it still seems that novelty and value alone already go a long way towards characterising episodes of creativity.

Additional or alternative criteria such as ‘surprising’ or ‘appropriate’ often refer to the character of the novelty or value as observed in the output. Even with ‘flair’, it is difficult to see how this property could be recognised or detected without recourse to the creative product, in which case it again comes to resemble a qualifier of novelty. For purposes of modelling, then, it seems reasonable to assume that these two characteristics, as observed to be expressed in the creative product, may be taken as diagnostic of a creative process.

However, lest it be thought that the ubiquity of the novelty and value identifiers imply homogeneity, it should be noted that both Boden and Kronfeldner distinguish varieties of creativity that depend on corresponding categories of novelty. Boden identifies further varieties which depend on different types of surprisingness. This range of variation will now be explored to determine whether further qualifiers will need to be added, or different categories of creativity will need to be explained (or taken into account in modelling).

In this, it will be convenient to start with the range of types of novelty, taken hand in hand with their corresponding classes of creativity.

2.3 Novelty

The introduction of anything new is an addition to the world, or to that part of it which forms the context against which the newness applies. Novelty being a basic criterion for creativity, any model designed to represent a creative process will have to include some mechanism capable of originating novelty.

It will be helpful first to look at the ontological background, and to establish some working assumptions as to how the world is made up, alongside discussion of the range of possible types of novelty (and thence also creativity). This section therefore begins by considering the nature of novelty itself, in particular such questions as: just what is novelty (in the current context), and what kinds of novelty might be possible, or achievable?

2.3.1 Nature and Art, or culture

Creative novelty applies to ideas, often as expressed in artefacts of one kind or another. Its scope is therefore normally limited to human thought and action, perhaps with extension to allow for artificial creativity, or rudimentary similar activity as might be observed in some animal species. Novelty in a more general sense occurs throughout the universe. Therefore it will be helpful to set up a cosmological scheme that allows

for distinction between creative novelty and other kinds. Mill's view of 'Nature' provides a convenient starting point:

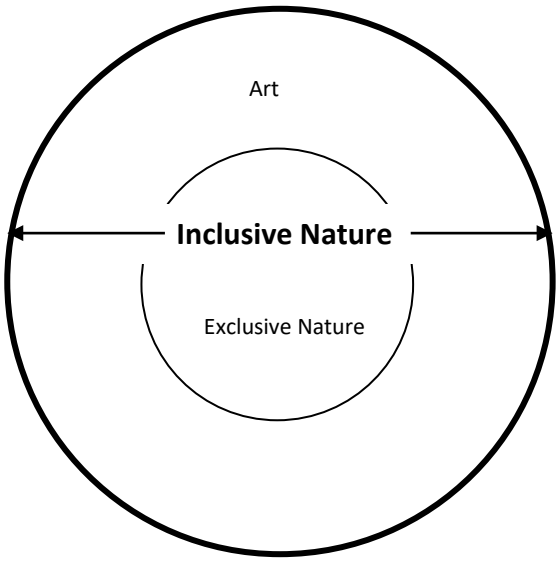
> In one sense, it [Nature] means all the powers existing in either the outer or the inner world and everything which takes place by means of those powers. In another sense, it means, not everything which happens, but only what takes place without the agency, or without the voluntary and intentional agency, of man. (1874, p.8).

The portion of the world which is covered by the first sense of Nature, but not the second (that is, all that is due to 'the voluntary and intentional agency of man'), he calls 'Art':

> [Nature in the first sense] entirely conflicts with the common form of speech by which Nature is opposed to Art, and natural to artificial. (1874, p.7).

This effectively divides the world into two parts along a convenient fault-line, separating, Nature in the second sense, (or exclusive Nature) and Art, these two combined going to make up Nature in the first sense (inclusive Nature). Inclusive Nature covers the entire cosmos, short of external intervention.

Fig. 2.1: Mill's Nature and Art

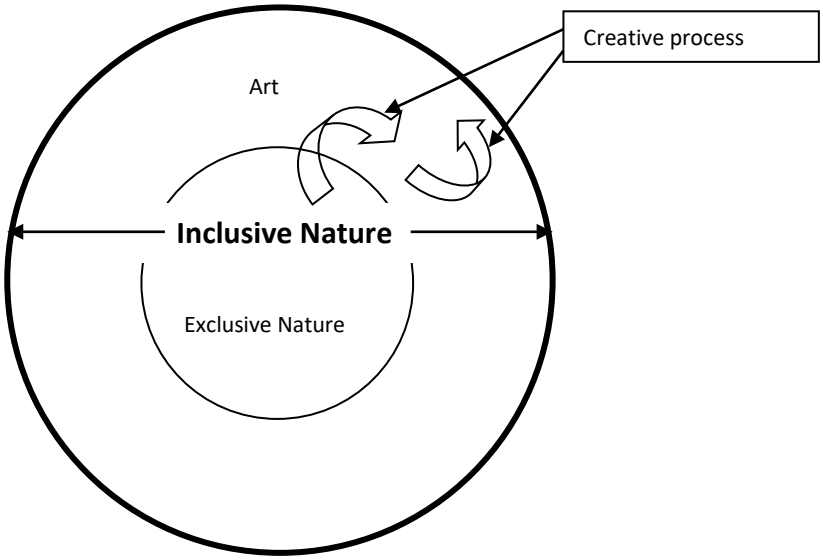


This schematic representation shows the division of Mill's all-encompassing inclusive Nature into Art and exclusive Nature (that part of the world that is free from the effects of human intentional action).

Creativity originates novelty in Art, not in exclusive Nature. This categorisation suggests an alternative way of describing creativity, aside from those based on novelty and added value, and that is that creativity consists in taking some part of exclusive Nature and transforming it into Art, or transforming some existing element of Art into a new form.

The curved arrows in the diagram below trace the effects of creative process in transforming elements of Art or inclusive nature into novel forms within Art.

Fig. 2.2: Creative process, relative to Mill's Nature and Art



Art, then, is composed of creative – or, at least novel – output (together with reproductions or copies of original works). A creative process adds novelty to Art, in a similar way to that by which the natural process of evolution might add novelty to exclusive Nature (although not necessarily via an identical or analogue process).

A particular instance of novelty, once having been added to Art, needs somehow to be preserved or it may be lost as the individual(s) involved in its production depart from the scene, resulting in the absence of any more permanent basis for further, cumulative, addition to culture. This points to a distinction between novelty in Art and in exclusive Nature, relating to the means by which the diversity of items within each is maintained. The differences between evolution and creativity will later be seen to be of some significance for understanding the relationship between the two processes and for the modelling exercise.

Kronfeldner uses ‘culture’ as a concept similar to (although not precisely coterminous with) Mill’s Art:

> Culture consists not only of art and entertainment, it includes domains of human activity of all kinds: science, religion, technology, language, etc. In general, culture includes all those activities that produce *cultural novelties*, i.e., entities which humans add to the world and which are maintained socially, i.e., are transmitted within and across generations by cultural tradition and not by biological heredity. (2009, pp.578-579)

Mill’s Art is not identical with culture because it takes more than a single individual to support a culture. Art allows for creative action by solitary individuals, and for creative output that is not communicated beyond the originator. Art is therefore more appropriate in a strict sense as a description of the sphere of operation of creativity but given that complete isolation is rare, culture can be considered for all present purposes as the sphere appropriate to creative novelty, as opposed to (exclusive) nature.

Therefore, for current purposes at least, it seems reasonable to accept that creativity refers to novelty in a particular context, that of the culture, and that one way of distinguishing culture is to recognise that it comprises that part of the world which is kept in being by social transmission.

2.3.2 Possible and actual

The above discussion of Mill’s view has taken Nature as referring to what is actual, whereas in fact Mill’s concept of inclusive Nature contains the notion that the world consists not only in what is (or once was) actually the case but also all that is possible, that is what might have been (counterfactual situations), or what could potentially come to pass:

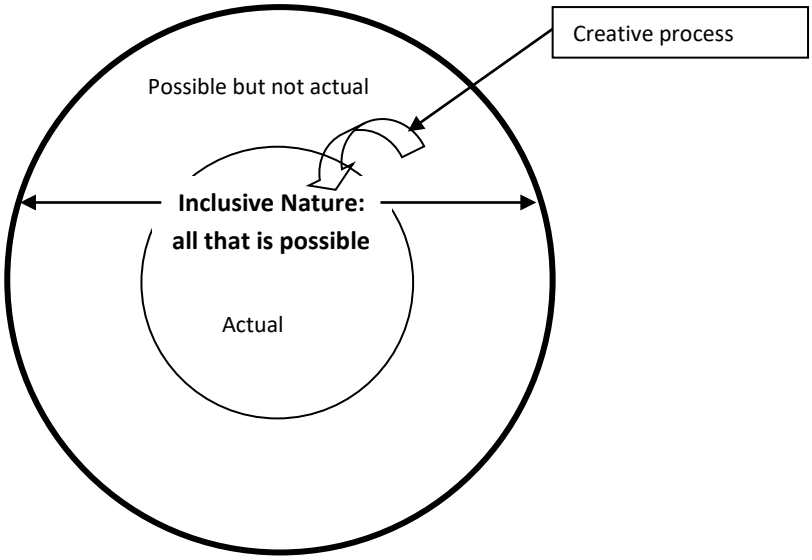
> Nature, then, in this its simplest acceptation, is a collective name for all facts, actual and possible… (1874, p.6).

If Nature is taken to embrace the possible, then all creative products are already included: those that have been, could have been, or are yet to be produced, or could be at any future time.

This suggests an alternative subdivision of inclusive Nature – all that is possible – with the actual world forming a subset. And then that in turn suggests a further description of creativity, in that it involves making actual something that was previously only possible, and at least partially unknown. (Biological evolution and other natural processes may also do this, of course, but not directly within the cultural sphere).

The following diagram, then, illustrates another way of representing a creative process; here the curved arrow traces the effects of creative process in rendering elements of the possible into actual forms of novelty.

Fig. 2.3: Creative process, relative to possible and actual



The entire creative action takes place within Mill's inclusive Nature, which takes in all that is possible.

Thus it seems clear that any creative process must access, and operate upon, the sphere of the possible, but not actual. The next section therefore considers the structure of possibility, and its implications for the nature of creative output (to which the basic criteria of novelty and added value apply).

Degrees (or grades) of possibility.

Mill's inclusive Nature can be taken as the universe of possibility, all that is possible under the most comprehensive interpretation. Then within this universe, a range of degrees of possibility – or 'grades', as Dennett (1995, p.104) calls them – can be distinguished in addition to the actual.

Of particular relevance to creativity is empirical possibility, that is, that which is possible under the actual laws of nature. Beyond (and including) empirical possibility lies the logically possible, which Dennett notes that '… according to philosophical tradition is simply a matter of being describable without contradiction' (1995, p.104) and thus true in at least one possible world. This category takes in situations which could hold if the laws of nature were different from the actual laws.

Beyond (and including) the logically possible there is that which is describable, whether or not it contains internal contradiction or inconsistency; Keith Frankish gives the example of a four-sided triangle (2005, p.49).

These degrees of possibility form a nested series similar to the one illustrated by Dennett (1995, p.107), who does not include the metaphysically impossible but adds further categories within the physically possible (historical and biological).

Any physical output of a creative process must be empirically possible, by definition, because it can only be formed as natural laws allow.

Within the empirically possible, there is further restriction on the scope of creativity, in that creativity, which entails the generation of novelty, and which in turn relates to present and future, cannot bring about counterfactual effects. Only physical novelty, achievable given the current actual situation, constitutes a potential result of a creative act.

This of course does not mean that counterfactual situations cannot be embodied in creative products, in the way that a historical novel or film may describe or depict events that might have happened but did not. The counterfactual events are

represented in physical forms which are themselves empirically possible. Similarly, of course, ideas that are only logically possible, or even logically impossible, can be communicated through empirically possible media.

Intuitively, the range of situations which can be imaged is more restricted than that of situations which can be described. Roy Sorensen even offered a prize:

> One hundred dollars to the first person who identifies a picture of a logical impossibility. (2002, p.337).

But for present purposes, it will be sufficient to note at this point that images seem to have a more restricted range within the universe of possibility (and impossibility) than do concepts and propositions.

To summarise, in terms of the scope of creativity, potential creative output thus forms a subset within the empirically possible – not counterfactual, and restricted within that to novel products which add value. The creative product may, of course, embody symbolic representations or descriptions of states of affairs that are only logically possible, or indeed impossible, but these can only be presented in empirically possible form. For example, a work of historical fiction may describe events that did not happen (or could not possibly have happened) in the actual world, but such events might be portrayed in the (itself empirically possible) form of a book (as text) or a film (as images). Similarly, no system within the actual world can produce output that is not empirically possible, and any modelled creative process will be subject to the same restriction.

Creative output, in addition to being empirically possible, must of course incorporate novelty. The following section therefore inquires into the range of possible categories of novelty, and their relation to creativity.

2.4 Varieties of novelty and their relation to creativity

As said, Boden and Kronfeldner refer to different types of creativity that in turn depend upon corresponding varieties of novelty. It is therefore convenient to consider the types of cultural novelty in question in conjunction with those of creativity itself.

2.4.1 Anthropological novelty and anthropological creativity

Kronfeldner's anthropological novelty is the broadest category of novelty, and applies to any (unoriginal) addition to culture (2007, p.169). This anthropological novelty is the output of an anthropological variety of creativity. 'Creativity' is here taken in a very broad sense, relaxing the novelty criterion by dropping any requirement for originality.

> To *create culture* means to bring about something that is new as a token. Making a traditional kind of pot is creating culture, even though it produces not more than a new token of a type of artefact that is already well established in a certain culture. This kind of creating is a *conditio humana* and a ubiquitous activity. That is why I call such activity *creativity in an anthropological sense*. (2011, p.36)

It is, however, difficult to see that there can be a simple one-to-one mapping from cases of anthropological novelty onto corresponding cases of anthropological creativity (as opposed to creation), for the reason that creativity generally carries the implication of added value (at least) in addition to novelty. To extend Kronfeldner's own example: manufacturing a pot to an established design may produce an item of value (as long as there is a demand for the product). But the continued production of a particular type of pot, in numbers greater than actually needed or desired, is less likely to be regarded as creative, even though novelty is still being generated (it becomes wasteful, and thus destructive, rather than creative). This distinction will be more pronounced for the historical and psychological varieties of novelty and creativity, which are required if any cultural development is to take place.

2.4.2 Historical and psychological varieties of novelty

Boden (2004, p2) and Kronfeldner (2011, p36) both distinguish historical and psychological types of creativity, which depend on categories of new ideas, or novelty; they each describe these categories in similar terms.

In contrast to anthropological novelty, the historical and psychological varieties do call for originality (relative or absolute). Kronfeldner delineates these types of novelty as follows:

> If the craftsman creates something culturally new (e.g., a pot), the product can be new in three further senses – either new for him (psychological novelty), or new for a group of people who are bound together diachronically and synchronically as a tradition (relative historical novelty), or even new in the sense of being its first appearance in the whole universe (objective historical novelty). (2009, p.580).

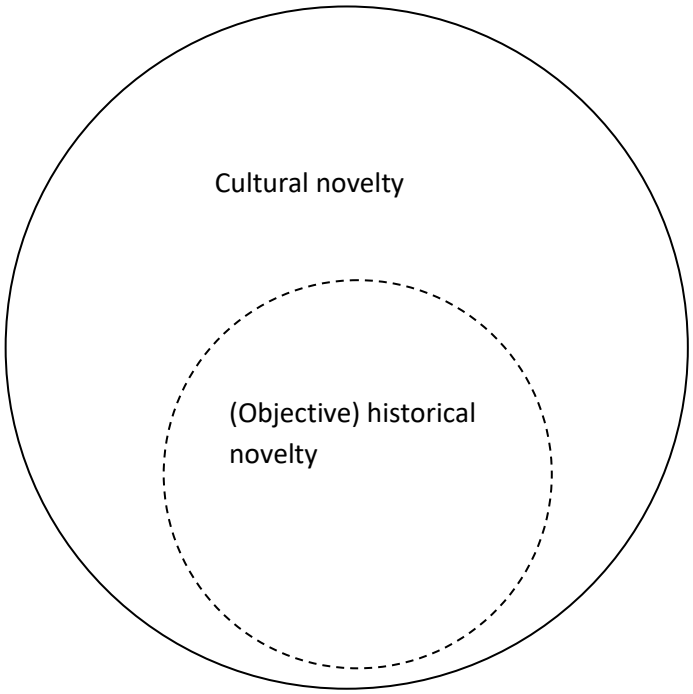
The historical variety is primary relative to the psychological, in that it refers to something completely new to the world, and thus potentially to an ontological event (at least as far as is known in the culture or society into which it is introduced). Something which is new to the individual (psychologically novel) need not be new to the culture in which the individual is embedded.

Thus in this context it makes a considerable difference whether novelty – or indeed creativity – is regarded from the microcosmic viewpoint of the individual, or from the macrocosmic one of the universe. Kronfeldner (2009, p.580) refers to these perspectives as ‘psychological’ and ‘historical’ respectively.

Of the four types of novelty Kronfeldner identifies, three form nested series, although the nesting varies according to the viewpoint. For the individual, something that is historically new is also psychologically new, as is anything which is learnt and/or copied (for the first time).

From the macrocosmic viewpoint (of the universe) only objective historical novelty, that which is completely new to the world, differs from other types of cultural novelty, of which it forms a separate subset:

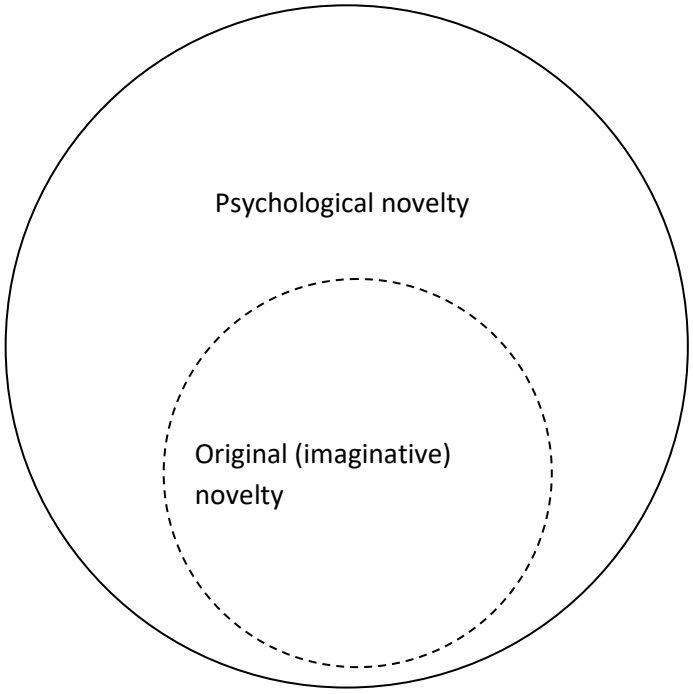
Fig. 2.4: Varieties of cultural novelty (macrocosmic viewpoint)



The diagram illustrates the distinction between historical novelty and all other instances of cultural novelty, taken from a universal perspective external to the creative agent. A pot of an entirely original design falls into that subset of cultural novelty which is historically new, whereas all subsequent copies made according to that same design will be new to the culture but will not constitute instances of historical novelty.

From the microcosmic viewpoint of the individual, the split is not in the same place; there are two types of psychological novelty, according to whether the source of the novelty is external (through copying or learning) or internal (from original imaginative thought).

Fig. 2.5: Varieties of cultural novelty (microcosmic viewpoint)



Here, original novelty is a subset within psychological novelty. Any historical novelty (whether relative or objective) produced by the individual is a further subset within original psychological novelty, but these are not distinct as far as the individual is concerned – the original psychological novelty is independently generated, whether or not it is also historically new. Thus, if a creative agent independently produces a new pot, to a design which they have not copied from any external original, then that is an example of psychological novelty. It may also be historically novel, provided there has been no previous appearance of that same design of which the agent was completely unaware.

2.5 Varieties of creativity

2.5.1 Historical and psychological varieties of creativity

As said above, cases of novelty do not bear a one-to-one relationship with creativity, because of the other factors involved in being creative, so that an instance of historical novelty does not necessarily imply historical creativity. For example, producing a lump of clay of complex but random shape might constitute historical novelty, but if the result had no artistic or practical merit it would not be creative in the way that designing and making a new type of pot might be (as the new shape fails to add any value).

But, all cases of creativity do require the generation of novelty of the corresponding type. The nesting relations between anthropological, psychological and historical novelty are mirrored in those for the equivalent forms of creativity, and depend similarly on the point of view.

> … If a potter has not copied the traditional way of making pots but just happens to make the same ones (unlikely but not impossible), then he might well be considered creative. What he does may be psychologically new, and in a psychological sense creative, even if what he does is historically old. (2009, p.580).

Psychological creativity requires the generation of novelty (which must be original, not copied or learnt (2011, p.36)) although not necessarily in the sense of historical originality.

Putting the microcosmic/macrocosmic discrepancy in a different way, Kronfeldner refers to it as an asymmetry (2009, pp.580-581), and points out that the psychological form is the more crucial from the point of view of the individual and the creative process. This is true even though from the macrocosmic angle it is only historical novelty that adds new ideas, entity types, or properties to the universe, and therefore carries ontological significance. And (as Kronfeldner says) when attempting to understand the origin of creative novelty, given that cultural novelty is generated by

humans, the psychological viewpoint is the more relevant because it is the generation of original psychological novelty that requires to be explained (2009, pp.580-581).

Microcosmically, there is no distinction provided that the individual is unaware of the existence of any originals. But, taking a macrocosmic view, there is a significant difference in that historically creative products add novelty to the culture, whereas other psychologically creative products do not.

Boden

Boden had distinguished psychological and historical varieties of creativity '(P-creativity and H-creativity, for short)' in *The Creative Mind*:

> P-creativity involves coming up with a surprising, valuable idea that's new *to the person who comes up with it*. It doesn't matter how many people have had that idea before. But if a new idea is H-creative, that means that (so far as we know) no one else has had it before: it has arisen for the first time in human history. (2004, p.2).

As can be seen from this extract, Boden brings in the value criterion, together with an additional requirement for surprise.

Boden had also noted the primary nature of historical creativity from the macrocosmic viewpoint, and of the psychological variety for understanding individual creative process (2004, p.2). And, Boden relates H- and P-creativity to corresponding types of novelty (as does Kronfeldner) (see 2004, p.43).

In sum, the first three types of novelty taken together afford a complete sketch of the ways in which nature is transformed into culture (always assuming the metaphysical variety is superfluous). Kronfeldner's (2009) identification of the anthropological varieties of novelty and creativity is a fruitful addition to the debate on creativity issues (the relationship between the psychological and historical types approximates to Boden's usage), even though anthropological creativity is not universally regarded as creative (lacking originality, surprise, and unexpectedness).

2.5.2 Metaphysical novelty and creativity

Kronfeldner describes one further concept of novelty, which (along with its associated form of creativity) she terms 'metaphysical' (although it must be borne in mind that she is herself arguing for a naturalistic explanation of creativity). Metaphysical creativity requires '*ex nihilo*' novelty (2009, p.582):

> According to Hausman ([1984] p.9), creativity by definition escapes from the causal fabric of the world. A creative thought or action is "unaccounted for by antecedents and available knowledge, and is thus disconnected with the past". (2009, p.582).

This is equivalent to inspiration or intervention in that the novelty appears *ex nihilo*, although here presumably no external source is necessarily assumed or invoked.

Three of the four types of creativity that Kronfeldner names form nested series as in the above diagrams of corresponding varieties of novelty. On the microcosmic view, historical creativity is a subset of the psychological form, and the anthropological variety is that portion of all creativity that generates output which is not psychologically novel with respect to the individual. But the metaphysical version of creativity does not sit easily within these series, even though Kronfeldner describes it as a narrower concept than any of the others: 'it sets such rigid conditions for creativity that creativity becomes impossible in a deterministic world.' (2009, p.582)

The other varieties all refer to characteristics of the novelty of the product or output, relative to the individual or cultural context. This fourth concept, on the other hand, has to do with the process by which the novelty is brought forth (undetermined by causal factors, and therefore free) and hence its source – as uncaused, it can only be derived *ex nihilo*, from nothing. Such events could no doubt give rise to instances of all three of the other types of novelty, the product being unspecified. For example, it is difficult to see why, if supernatural muses are at work, they might not prompt cases of psychological novelty as readily as the historical variety (in the ignorance of the individual concerned regarding the prior existence of an original).

Thus, although the boundaries of this category are tightly drawn as regards the process and source, it is unrestricted in terms of the product and its relation to context, so even if the frequency of such cases might be low, and the definition narrow, its potential scope of operation appears to be so wide as to encompass the entire range of psychological novelty.

It would seem to be virtually impossible to establish that metaphysical creativity *never* happens (or that no creative idea is ever introduced into an individual's mind by supernatural agency), but what Kronfeldner (2009) aims to demonstrate is that all forms of creativity are compatible with determinism, such that there is no need to call upon metaphysical theories of creativity (which would presumably then fall to Occam's razor). In that case, creative novelty could be fully accounted for by reference to prior states of affairs, and would thus be explainable in those terms.

2.5.3 Causation, determinism, and ex nihilo *creativity*

Determinism itself is a complex topic, on which space and time will not allow discussion in depth. For current purposes, it will be adequate to regard determinism in a straightforward, classical sense as regarding all events to be, in principle, explainable in terms of cause and effect, according to laws of nature. In contrast, any effects arising from intervention by forces external to nature, or otherwise *ex nihilo*, would not support explanation. Nor could any process be understood or modelled if such influences were at work. On the other hand, if a model could be constructed for a process which would be capable of generating output satisfying criteria for a creative product, then that would demonstrate that creativity in the absence of any *ex nihilo* component was at least conceivable (although it would not prove that human creativity worked in the same way),

This approach ignores the problems associated with determinism, of course. For example, the position becomes much less clear at molecular, atomic, and smaller levels of magnitude, where quantum effects might begin to become significant. If

creativity as a psychological phenomenon is regarded as ultimately grounded in neural processing, then quantum mechanical considerations may have to be taken into account in any attempt to understand its operation. However, this would not in any case be achievable with present day technology. Therefore a higher-level process is sought which corresponds to the production of creative novelty, in much the same way that natural selection describes and explains evolution at a biological level. (It might be added that biological evolution also involves molecular-level events, in the replication of DNA, for example.)

Determinism links to questions of the freedom of the will, but the type of process sought would depend on cause and effect, and would thus be independent of the will in terms of its internal operation (much as for the processing within a Fodorian input module, as described in the next chapter). It might also be recalled that Kant had noted that creativity does not appear to perform at the behest of the will (see chapter one).

Kronfeldner (2009) attempts to set up a naturalistic account, compatible with determinism. She argues that determinism allows for change, and therefore cannot exclude novelty (2009, p.583) (although ancient and medieval cosmologies –admittedly less deterministic – did allow for change, but not generally for evolution, so that even if novelty was not totally excluded, its scope was very much restricted, even though change was not forbidden).

Kronfeldner also declares that ‘determinism does not imply that an observable change is already entailed in the antecedent conditions’ (2009, p.583). At first sight, this seems an odd claim, as on the face of it that is just what determinism does imply. Perhaps entailment only applies given complete knowledge of *all* antecedent conditions (i.e. the full configuration of the entire universe, and thus unattainable in practice). Such an approach would allow for Kronfeldner’s distinction (2009, p.583) between the entailed possibility of an event and the (non-entailed) actualisation of that same event. This

would provide for the effect of unknown antecedents, as in Kronfeldner's example (2009, p.587) of the coincidental drop of a loose stone, representing an influence emanating from beyond the problem space.

Problem spaces, ex nihilo *solutions, coincidence, and serendipity*

It seems doubtful whether Kronfeldner can really equate an influence emanating from outside the problem space with something arising *ex nihilo*, notwithstanding it may be unexpected or surprising. In Kekulé's case, for example, knowledge of the properties of snakes would have been present as a learnt antecedent (even if in an unrelated domain of thinking). (Also, loose stones do not generally appear *ex nihilo*, even given that they are unanticipated; there are reasons why they fall, and they have their own causes that might be found in the weathering of fixings, for example, albeit that those causes might lie outside the problem space currently under consideration.)

Similarly, the relationship with coincidence also seems problematic. Relating contents extracted from two otherwise separate domains of thought (organic chemistry and snakes) does not necessarily count as coincidence (Fleming's discovery of penicillin might be a different matter – it more closely resembles the loose stone example – although this might perhaps be more precisely described as involving serendipity rather than coincidence). In fact it seems such ideas can be sourced from anywhere outside the problem space, whether in the external world (coincidence/serendipity) or in memory (through association).

Creativity and causal independence

Kronfeldner argues that creativity only demands independence from certain specific causes: learning and experience. She claims that in order to count as creative, psychological novelty must transcend them both.

Originality clearly does require independence from learning and experience, exactly as claimed, but only from their usage in direct or derivative fashion. Kekulé needed to

know about snakes and organic chemistry in order to gain his insight into the structure of benzene. Kekulé's example does not so much transcend learning and experience, as use them in an indirect way (which is not derivative, or routine). In fact, (as Kronfeldner notes) Kekulé's prior knowledge was necessary to formulate the idea, and to recognise its significance (an example of the application of the value criterion), although that is not the same as prior knowledge directly leading to a solution, as in the routine drawing on learning or experience to handle everyday situations, or to solve familiar types of problem.

There would, nevertheless, be a sense in which (previously only latent) novel properties are not reducible to antecedent states, if prior knowledge could not predict their emergence (this might then go some way to support the element of surprise that Boden recognises in cases of creativity).

The construction of a mechanical wristwatch is entailed given:

1. The necessary components;

2. A process in assembling them (requiring instructions).

The property of telling the time is not part of either the components or the process, even though it arises out of them. It is a question as to how far the new *property* (as opposed to the new configuration of the components in the finished article) is entailed by either (1) or (2), or both. How easily is it explained by reference to them? In a newly-created product, the properties might not be known, or not fully anticipated, prior to the creative action. This may lead to a conclusion to an extent similar to Hausman's but without calling for metaphysical freedom from any causal chain.

2.5.4 Varieties of creativity relative to surprisingness

As said above, Boden ascribes the qualities of being *novel, surprising, and valuable* to creative output. Besides P-creativity and H-creativity, which refer to different types of

novelty, Boden identifies three varieties of creativity corresponding to alternative ways in which novelty can be surprising. These varieties are, respectively: combinational, exploratory, and transformational types of creativity. Of particular interest in the current context (given the aim of constructing a model creative process) is Boden's assignment of distinct mechanisms to each of the three varieties.

> These three forms of surprise correspond to three types of creativity generated by different psychological mechanisms (Boden 2004). The first is *combinational* creativity, which involves the unfamiliar juxtaposition of familiar ideas. (Think of poetic imagery, musical "quotations," or visual collage.) The generative processes here are those of associative memory. (2014, p.228).

Boden's first way of being new and surprising, combinational creativity, is the antithesis of Kronfeldner's description of the concept of metaphysical creativity, in that there is nothing obscure about the source (already familiar ideas), and no fracture in the chain of causality is needed to account for it. The novelty arises purely from a rearrangement of pre-existing elements. In this, it resembles novelty generated in nature by evolution through natural selection, where no new physical components are required, but novel properties arise from new combinations of already existing physical entities.

(It might be noted here that the requirement for relevant prior knowledge recalls Wallas's preparation stage. The problem of identifying appropriate combinations from among the many that might be possible will be discussed later on.)

Boden's exploratory and transformational creativity, and conceptual spaces

Exploratory and transformational creativity, Boden relates to conceptual spaces, which are the 'structured styles of thought', accepted within a culture (2004, p.4). Boden provides as examples:

> …ways of writing prose or poetry; styles of sculpture, painting or music; theories in chemistry or biology; fashions in couture or choreography, *nouvelle cuisine* and good old meat and two veg – in short, any disciplined way of thinking that is familiar to (and valued by) a certain social group. (2004, p.4).

Boden also includes the permissible moves within the rules of games such as noughts and crosses and chess (2004, p4). She notes the vastly greater number possible for chess, and says that:

> Whatever the size of the space, someone who comes up with a new idea within that thinking style is being creative in the second, exploratory, sense. (2004, p4).

It seems clear from this that a conceptual space is a sector of the universe of possibility, rather than of the actual world. The extent of the relevant space expands rapidly as the rules defining it go from those for the relatively simple noughts and crosses and on to the more complex ones for chess. Its range is determined by a corresponding 'generative system' (such as the rules of grammar (1994 p.78), for example), which marks out its boundaries.

Thus, a conceptual space spans all the possibilities allowed within the rules which specify it (the generative system). In this, it resembles a *possibility space* in statistics, defined as follows:

> The set of possible outcomes of a probabilistic experiment is called a possibility space. For example, if two coins are tossed the possibility space is the set of possible results HH, HT, TH, TT, where H indicates a head and T a tail. (Hastings and Peacock, 1974, p.2)

The possibility space is also a sector of the universe of possibility, which itself is just the set of all points in all possibility space.

No assumption is intended here as to the ontological status of the universe of possibility – it is proposed purely as a construct for the sake of argument, without any position being taken as to its objective existence, or non-existence, but nonetheless treating it as if it were an objective ontological fact. The purpose behind this is not to make any strong metaphysical claims, but solely to construct a framework for further discussion; it can therefore be discarded after use.

Boden's exploratory creativity involves generating a new idea within already established conceptual space, and according to the accepted rules of its generative system.

Boden bases the distinction between exploratory and transformational (or radical) creativity on the relationship between the novelty originated and the appropriate conceptual space or generative system (1994, pp.78-79). The conceptual space defines a range of possibilities, and she says that anything outside that range is '*impossible'* (1994, p.79) (that is, within the constraints imposed by the generative system). However, a conceptual space can be transformed by expanding its boundary within possibility space (1994, pp.80-81).

Such transformation may be achieved by negating or dropping constraints on the conceptual space:

> Dropping a constraint is a general heuristic, or method, for transforming conceptual spaces. The deeper the generative role of the constraint in the system concerned, the greater the transformation of the space.
>
> Non-Euclidean geometry, for instance, resulted from dropping Euclid's fifth axiom, about parallel lines meeting at infinity. […]
>
> Another very general way of transforming conceptual spaces is to "consider the negative": that is, to negate a constraint. (Negating a constraint is not the same as dropping it. Suppose someone gets bored with eating only red sweets: to choose *any nonred sweet* is different from choosing *any sweet, whatever its color*.) (1994, p.82).

Boden (1994, pp.82-83) gives Kekulé's discovery as one example of transformational creativity via considering the negative (that carbon atoms might not always form open chains). And yet this could perhaps be taken as combinatorial – the combination of pre-existing open chain theory with pre-existing ideas about snakes, and flexible and closed chains. Also, for the negation or dropping of constraints it might be necessary to already possess some understanding of the nature of the constraint (cf. Wallas's preparation stage), maybe together with the general idea of negation (or nullifying).

To summarise Boden's three types of creativity relative to surprisingness: combinational creativity refers to the source of the new ideas (namely, the already existing ideas which are subsequently recombined) whereas the exploratory and transformational varieties refer to the creative process and the location of its output in possibility space, relative to existing ideas in the same field.

Boden's exploratory and transformational varieties provide an important route for approaching creativity, and also of understanding the creative process, particularly from the point of view of the possibility spaces concerned.

2.6 Potential sources for creative ideas

Boden's combinational creativity, which results from new combinations of pre-existing ideas, raises the question of the raw material out of which the creative output is formed. There are intuitively only three potential sources for new ideas:

1. Internal source: they are composed of elements of existing ideas, memories, perceptions, thoughts, concepts, or other prior mental content, as for Boden's combinatorial creativity (here taking Boden's established generative systems as pre-existing notions).

2. External source: they are found in the outside world (as in serendipitous discovery). Or, perhaps more accurately, they are taken from mental content formed in response to the external world (that is perceptual data, including communications from other individuals, or in Fodorian terms, the output from perceptual and language modules).

3. They come from somewhere else, which can only be nowhere, in other words *ex nihilo* and presumably from beyond the reach of any possibility space. (This would correspond to Kronfeldner's description of metaphysical creativity).

(Of course, new ideas might be sourced from a mixture of any of these three.)

To take a naturalistic view of creativity (implying that only sources ① & ② are valid) is to deny the reality of Kronfeldner's metaphysical variety. This is tantamount to assuming that the universe of possibility resembles a universal conceptual space corresponding to a universal generative system (the laws of nature). This contains all possibility space, and there is no influx of new possibilities across its boundary; hence the scope of creativity is constrained by that boundary. In other words, universal radical novelty (using Boden's terms), or metaphysical creativity (Kronfeldner's concept), is impossible.

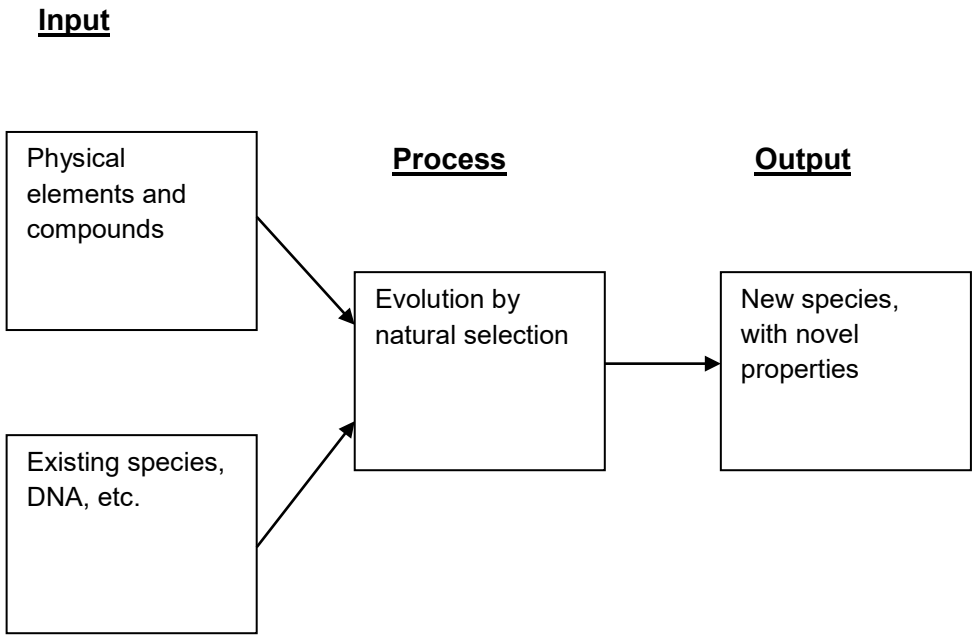
Given this assumption, and turning to what it implies for the source of novelty, as opposed to the processes by which it is generated, we have the following.

Input sources for the generation of novelty in evolution

In nature, evolution by natural selection generates novelty in forms of life, but no metaphysical source is needed to account for this: all forms of life are built of the same chemical elements: carbon, hydrogen, oxygen, nitrogen, and a few more. These are therefore the material sources, in the sense of pre-existing components. They are arranged and rearranged into many complex forms and structures: DNA, proteins, cells, molecules, nerves, muscles, sense organs and so on, depending on species. The rearrangements occur naturally and organisms with novel properties, such as: photosynthesis, locomotion, sensation, and cognition, have emerged over time through the evolutionary process. But regardless of the type of process, and the emergent properties, the physical ground of life in all its varieties remains the chemical elements present in the earth (no new elements having been synthesised until very recently, and those that have, have not been incorporated into living structures or metabolisms).

Thus new arrangements of physical antecedents allow properties hitherto latent in the universe of possibility, to emerge into actuality.

Fig. 2.6: Evolution by natural selection: inputs and outputs



In the above sketch, the process of evolution by natural selection is depicted as transforming the (physical) input materials into outputs which consist of successful new variants. Any model for a creative process constructed on similar lines would need to take account of the equivalent input sources, and thus the next section gives further consideration to this aspect.

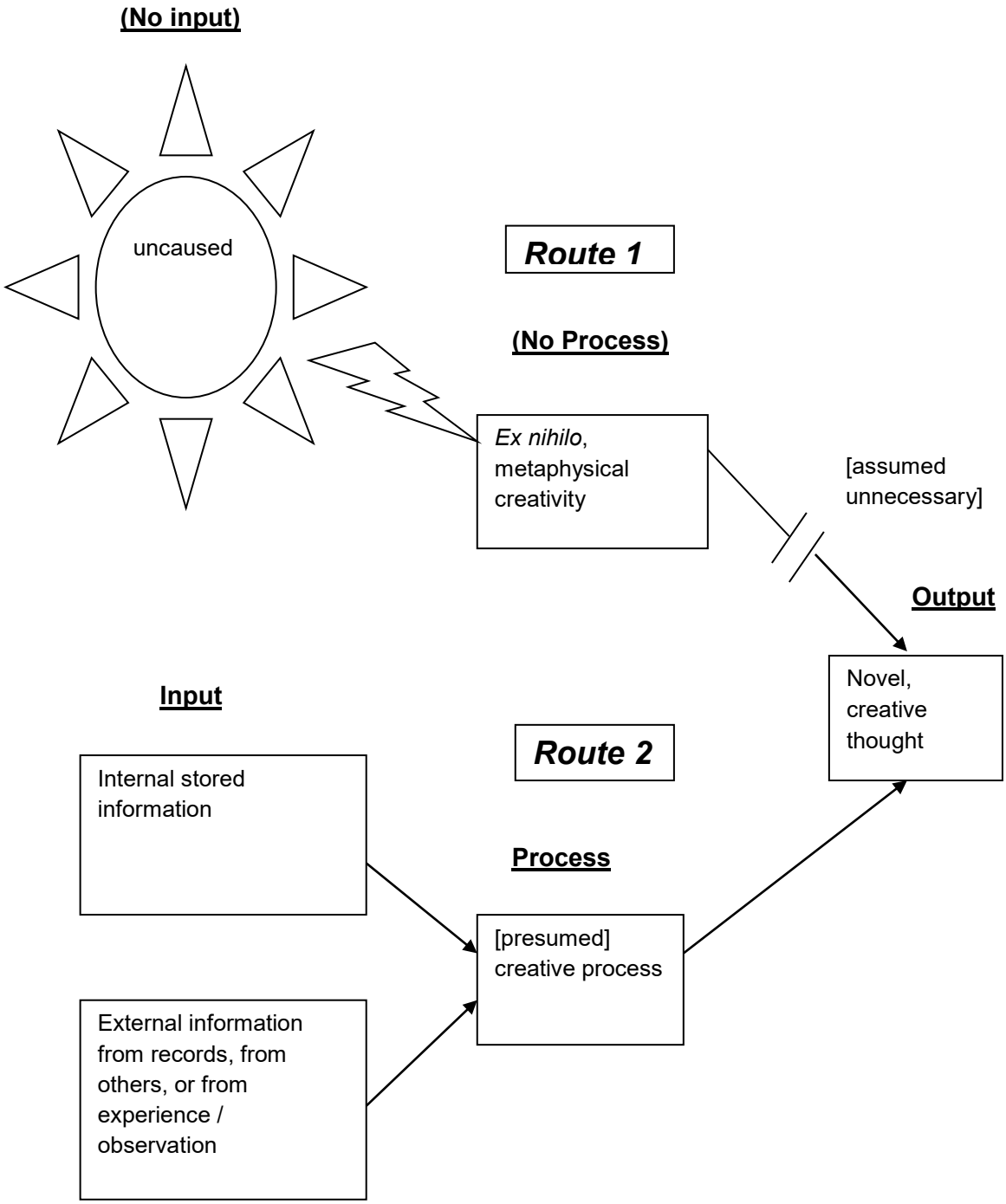
Input sources for the generation of novelty in creativity

In culture, novelty is generated via creativity and creative process. If historical novelty, as an ontological event, is produced within the laws of nature, and is traceable to antecedent causes, then the source (in the sense of *input*) must be within the universe, and, as said, its elements must be pre-existing ideas, subcomponents of them, information, or observations. A creative individual might draw such material from

within, from stored information or data in memory, or from representations or images. The only other potential source is whatever is external (and accessible) to the individual, in the environment, either as information already available in the culture (from artefacts, recorded material, or from communications from others), or from observation or experience of the external world, as in serendipitous discovery, for example.

The potential sources of input to a creative process are summarised in the following diagram.

Fig. 2.7: The creative generation of cultural novelty – potential input sources



Here, route 1 corresponds to metaphysical creativity, and route 2 to all other varieties. Route 2 can be taken as any traceable process (regardless of the particular mechanism involved) leading from the conjectured input sources to creative output. The output, novel creative thought, will include some points in possibility space previously unknown to the individual (in psychological creativity), and perhaps also new to the whole culture (historical creativity).

In route 1, the output arrives *ex nihilo*, out of nothing (or, from beyond the boundary of the material cosmos and the actual world). Therefore there is no input, or alternatively, it cannot be asked what items were used to make the product. Nor can it be asked how (by what process) the input was formed into the output; the product arrives in the absence of these antecedents. Such events could not be explained or understood.

In contrast, for any non-metaphysical form of creativity, there is no *ex nihilo* element, and so it *can* be asked how the output was produced, and how the novelty was originated. Furthermore, if a viable process can be specified that is plausibly capable of generating creative novelty, then that will establish that it is conceivable that such novelty could be originated via route 2. And, provided that creative novelty can potentially be explained in terms of route 2, route 1 falls to Occam's razor; it becomes unnecessary to invoke metaphysical origins. This, then, is what the modelling exercise will aim to achieve.

2.7 Creative process

A process is here taken to consist of the operations that need to be carried out to reach an end point, product, or output. If a specified set of steps could be shown to lead to creative novelty, then that would explain how creative products might possibly be achieved, without recourse to any *ex nihilo* element (in the same sense in which the process of evolution by natural selection explains the origin of biological novelty).

This section therefore moves on from considering the sources of input to creativity, and begins to focus attention on the creative process itself; that which is required to operate upon the input material in order to generate the creative novelty. It will be convenient to commence by examining the requirements for such a process – that is, for the task which it has to perform. This task could be described as taking something previously unknown from the universe of possibility and rendering it actual (if only in thought). The first questions to be addressed relate to the universe of possibility: what may be found within it, and how might the appropriate content be located? Or, in other words, what are the characteristics of its contents, and how might they be accessed or observed? These issues will be approached by way of a metaphor, in the following sections, but first it is necessary to look at the relation between the universe of possibility and the actual world, including the relative extent of each.

2.7.1 Possibility and actuality

Vlatko Vedral, in *Decoding Reality*, argues '... that information (and not matter or energy or love) is the building block on which everything is constructed.' (2010, p10) He estimates the maximum number of binary digits, or bits, of information that can be accommodated by the actual, physical universe by means of a formula devised by Jacob Bekenstein, whose:

> ... relationship, known as the Bekenstein bound, is simply stated as follows: the number of bits that can be packed into any system is at most 10^{44} bits of information times the system's mass in kilograms and its maximum length in metres (the square of this length is the system's area). (2010, p.185)

Vedral takes the following as estimates of the extent and mass of the universe:

> Astronomers have already given us a rough estimate on the Universe's size and weight, say 15 billion light-years in diameter and a mass of about 10 to the power of 42 kilograms ... (2010, p.186).

So that the Bekenstein figure for the limit of the information holding '...capacity of the universe ends up being on the order of 10^{100} bits of information. This is a stupendously large number, but ultimately it is not infinite.' (2010, p.186).

The psychological dimension

Vedral also employs the same formula to estimate the overall information holding capacity of the human head:

> Say that a typical head is 20 centimetres in diameter and weighs 5 kilograms. That means that a typical human head can store 10 to the power of 44 bits of information. Compare this to the best current computers which are still only of the order of 10 to the power of 14 bits of information. We therefore need 10^{30} to get the equivalent information carrying ability of a human head! (2010, pp.185-186).

Presumably the capacity of those components of the brain which are capable of storing information used in, or accessible to, mental processing and memory would be less than this, but the comparison with computing is nonetheless apt.

Thus the overall potential for forming combinations and recombinations of the elements of information held by an individual, while many orders of magnitude less than that for the universe as a whole, and whether taken on the basis of head, brain, or mind, must still be correspondingly vast, at least in theory. This is of some significance if creativity is regarded as a psychological process (and particularly so for combinatorial theories of creativity).

An easy way to approach looking at the relative extent of the universe of possibility and the actual world is to take a subset from a fictional source, which will repay examination in some detail.

2.7.2 The Universal, or Total, Library of Babel

The Library of Babel is the title of a short story by Jorge Luis Borges, in which he explores the relationship between the possible and the actual, along with some of the problems in working with the possible.

The story concerns a fictional library which holds books corresponding to every combination of a specified set of alphabetic characters and punctuation:

> ... all the books, no matter how diverse they might be, are made up of the same elements: the space, the period, the comma, the twenty-two letters of the alphabet. (1942, p.81).

The books and their constituent pages are of uniform, standard lengths such that each book contains 1,312,000 characters.

The *Library of Babel* is, in effect, the set of all possible books because such a library would include somewhere within it every book that could be written using the specified set of symbols (books of more or less than the standard length could be accommodated as part volumes, or spread across more than one volume as appropriate, but see below). It would of course be limited to combinations of symbols and so does not include entities or events, although these would be described somewhere within its contents. The library is therefore the complete set of outcomes, or the entire possibility space, corresponding to a specified generative system.

The books do not appear to be arranged in any particular order, and there is no index. This randomness, together with the inevitable preponderance of unintelligible text, renders the *Library* unusable.

Because of these limitations, the *Library* is of no more value than having no books at all. As a random string of symbols, it contains huge quantities of data but no information (taking information to be data that has meaning). Russell K. Standish notes that:

> Far from containing the wisdom of the ages, the library is useless, containing no information of worth. Our libraries are useful, not so much for the books they contain, but for the books they don't contain! (2011, locn.255, 5%).

This is because there is nothing to indicate which from among its vast number of volumes belong to the vanishingly small proportion that contains meaningful text (Borges covers this point very well in the original short story). Thus it is as if the *Library of Babel* is both (a) all possible information, and (b) a complete absence of information, at one and the same time. Standish makes a similar point:

> There is a mathematical equivalence between the Everything, as represented by this collection of all possible descriptions and Nothing, a state of no information. That some of the descriptions must describe conscious observers who obviously observe something, gives us a mechanism for getting Something from Nothing: *Something* is the “sw" of *Nothing*. (2011, locn.130, 2%).

The collections in actual libraries are all subsets of that in the *Library of Babel*; indeed an actual library is only of any value for the very reason that it is a selected subset. And given all this, the task (or process) of writing and publishing an actual book might be taken as in some sense equivalent to that of selecting the corresponding volume from the *Library of Babel* and placing it in the set held within all actual libraries (and viewed in this context, the process even appears to have something in common with the generation of creative novelty from nothing, *ex nihilo*).

The *Library of Babel* corresponds to all possible books (within the set limits), this being a subset of the sum of all possible entities and events, over all possible worlds. It can thus be taken as a proxy for the universe of possibility.

The overall size of the Universal Library

Borges anticipated his short story in his 1939 essay entitled *The Total Library*, in which he traces the origins of the basic principle underpinning the *Library* back to Aristotle[24] (1939, p.214). The concept of the *Library* as such Borges attributes to Kurd Lasswitz (and before that to Gustav Theodor Fechner).

Lasswitz published his own, earlier, short story on the same theme (entitled *The Universal Library*) in 1901. In this, he includes a calculation of the extent of the Library, which works out at $10^{2,000,000}$ volumes[25]. This number is $10^{1,999,900}$ times greater that Vedral’s estimate for the information holding capacity of the physical universe, and $10^{1,999,900}$ would itself of course be an underestimate of the minimum number of such universes required to hold the data in the *Universal Library*, because every (Lasswitz)

[24] Presumably including the *Metaphysics* (1991, 985b, p.22).
[25] See: Lasswitz (1901, p241).

volume is composed of 1,000,000 symbols, each of which would require more than one bit to encode.

The sizes of (a) the character set, and (b) each volume, chosen to form the basis of the Library determine its overall extent. Smaller volumes would still cover the same material, but there would be correspondingly greater use of multi-volume works, of which the constituent books might be widely separated. Willard Van Orman Quine takes up this point in his brief note entitled *Universal Library* (1990, pp.223-225). He bases his own initial calculation of the overall number of volumes on the same page size as Lasswitz, but takes a set of 80 characters and volume length of 250 pages. Quine proceeds to simplify the makeup of the *Library* by adopting a version of Morse code, consisting of two basic characters, and thus reducing the number of volumes to $2^{500,000}$ (retaining the same page and volume lengths). Each volume contains less information than the 80 character version, '...but there is no loss in content over all, since for each cliff-hanging volume there is still every conceivable sequel on some shelf or other.' (1990, p.224).

The reduced form of Morse is of course very similar to binary code, and this combined with the retention of content on cutting down the volume length leads to Quine suggesting an ultimate abbreviation to two volumes, each containing one of the only two symbols in a binary character set. Thus the two ultimate constituents of any piece of text, or information, themselves in a sense correspond to both no information, and all possible information, at one and the same time, just as does any complete listing of the data corresponding to all possible combinations of any standard number of those characters, as in the *Universal Library*.

> The miracle of the finite but universal library is a mere inflation of the miracle of binary notation: everything worth saying, and everything else as well, can be said with two characters. (1990, p225)

The* Universal Library*, random generation of text, and creativity

These two fundamental constituents can of course be comfortably accommodated within virtually any subdivision of the physical universe, so that an alternative approach to constructing parts of the *Library* is to arrange to produce random sequences of symbols. Borges himself alludes to this, quoting one of the speakers in Cicero's *De Natura Deorum* (1939, pp.214-215). He also mentions a more modern version of random text creation:

> Huxley ... does not say that the "golden characters" would finally compose a Latin verse if they were thrown a sufficient number of times; he says that a half-dozen monkeys provided with typewriters would. In a few eternities, produce all the books in the British Museum. (1939, p.215, footnote omitted).

Such random methods of text construction are very capable of producing completely new arrangements of symbols, thus meeting the first (novelty) criterion for creativity, but the likelihood of the output having any worth is vanishingly small, so that they will inevitably fail to satisfy the second (value) criterion. This observation will need to be borne in mind when theories of creativity involving some randomising element are considered. It may be a particularly significant factor when investigating approaches based on an analogy with evolution, and calling for the generation of novelty through a process involving any random, semi-random, or combinatorial elements.

Nevertheless, there obviously are some combinations of letters (otherwise not in any way special) that give rise to emergent properties, in that they constitute meaningful books. They thus correspond to cultural items (and, if not previously encountered, novelty) embodying cultural value. The difficulties with such random methods of generation are to do with targeting such fruitful combinations, and not in any question as to the existence of the combinations themselves.

Even in the (vanishingly improbable) event that a new novel, say, came to be produced by randomly combining letters and symbols, it could not be counted as an instance of creativity, in spite of the evident origination of novelty and value (this is exactly as required by Gaut – see earlier in this chapter). The novel would not be the output of a

creative process, in much the same way that a new species of microbe that spontaneously appeared in a mixture of organic materials by random recombination of molecules would not be a product of evolution. Similarly, the acquisition of a large sum of money by the chance purchase of a winning lottery ticket could not be considered to be the result of successful business enterprise.

In sum, the random production of novel combinations is clearly hopelessly inefficient on its own as a method for seeking any meaningful (or value-adding) result, so how might such an end be attained or brought about through creative (or indeed evolutionary) processes, given such insuperable odds against such an outcome?

2.8 The *Meno* problem

The question posed by the *Meno* paradox was introduced in chapter one. It asks how novel solutions can be achieved when: if it is known what is looked for, then the result cannot be new, but if it is not known, then how can a search be conducted?

The paradox presents as a dilemma, as Nickles (2003, p.56) points out, one of the horns of which boils down to a matter of recognition:

> … if you do not know the answer, then you also cannot inquire, since you would not *recognize* the answer even should you stumble upon it accidentally. Ability to recognize is the key, and you have to know an item already in order to recognize it. (2003, p.56).

How is it then that creative, meaningful novelty continually appears in the actual world, in spite of Socrates' dilemma? This issue will be taken up in chapter five, but for the moment perhaps an analogue of recognition, even though it is itself a mental event, might be observed to arise from environmental challenge in evolution. For example, if the climate in the habitat of a particular species becomes drier, then those genetic variants with better tolerance of drought are likely to be favoured. The increase of such variants is brought about by the change in climatic conditions, but this might alternatively be viewed as their being (analogue) 'recognised' as adaptive by the

environmental niche thus opened up. The ‘recognition’ arises from the fit of the properties of the successful variants to the environmental constraints.

Nickles offers a generalised trial and error process as a route to resolving the dilemma.

> As far as I know, the only viable solution involves an element of trial and error, that is, a process of blind variation plus selective retention (BV+SR). (2003, p.56).

As Borges and others have demonstrated, this on its own will not suffice, because of the improbability of random trials yielding an appropriate result. (Just as passing electrical discharges through a mixture of simple compounds of carbon, hydrogen, oxygen, nitrogen, and other constituent elements of living things will not produce even a single-celled bacterium).

The selective retention component allows for changes in evolution to be cumulative, as Nickles says.

> … things are very different if the standard of comparison is greater fitness relative to other members of an extant population, if the fitter individuals have a greater chance of selection, and if the retention mechanism involves transmission, with variation, to a new generation such that we get iterated ‘cycles’ of BV+SR. In that case, we get gradual evolution. Darwin was the first to recognize this, and he further recognized that a BV+SR process, when it is in place, does not merely make evolution *possible*, it makes it (nomically) *necessary*, assuming a reasonably stable environment with selection pressures. (2003, p.56).

One further point about innovation through random, combinatorial methods is that it is possible to get out more than was put in. If a random combination of alphabetic characters happened to spell out the text of Shakespeare’s *Macbeth*, then this would have many properties absent both from the original component symbols, and also from the vast majority of other such outcomes of random trials. These latent, emergent properties reflect added value, something over and above the sum of all the source material. It is difficult to see how in these circumstances the appearance of the extra properties could be predictable, hence the surprisingness. The chance generation of an entirely new work of literature by the same method would presumably bypass the

Meno problem if there were a current demand for the product, in which case it might be recognised as having value.

Following Dennett (1995, e.g. pp.68-73), Nickles (2003, p.59) refers to the combinations supporting useful emergent properties as embodying novel *design.* In such a case, the input raw materials were of course in existence prior to their reconfiguration, as Nickles notes (2003, p.59), but they did not at that time exhibit the new properties deriving from the pattern or design, which itself was not there beforehand.

Nickles mentions random mutations (2003, p.59) as an input source for BV + SR. These are of course also a rearrangement of pre-existing matter, although taking it into combinations that would not otherwise have arisen given the initial configuration (which would have been according to some previous pattern or design). In this, mutations might perhaps be taken to resemble serendipitous discovery, where there is some external influence, rather than any human creative process drawing purely on internal (mental) resources.

Also, newly emergent properties are present neither in the environment nor in any of the constituent physical matter, but are latent until the 'right' combinations are found. In a footnote, Nickles says that:

> Useful though this is, talk of search spaces and design spaces, as if all the possibilities already existed there waiting for us to discover them, can be misleadingly Platonic. It also threatens to reduce all discovery or recognition to the Columbus-discovers-a-new continent model. […] (2003, p.59 footnote 10)

But, emergent properties effectively lie outside the search space, and in some way beyond the set of possible output combinations alone. Rather, it is as if a new dimension of possibility space is opened up, and thus more as though Columbus had instead found a portal to a different zone of the cosmos, in a manner perhaps recalling Boden's transformational variety of creativity.

As Nickles points out (2003, p.59), the emergent properties, previously absent from the constituent matter, effectively arise *ex nihilo*, albeit in the context of wholly pre-existing input material. The new configuration reduces to its pre-existing components, but the embedded design does not (2003, p.59). In that sense, perhaps there is an element of discovery (in this case of new, hitherto unsuspected dimensions of possibility space). But discovery goes hand in hand with the recognition of value to which Nickles has drawn attention. That recognition requires the presence of a known requirement for a particular type of value (analogous – in the context of natural selection – to the existence of a suitable environmental niche, as noted above). To put this in terms of the *Meno* paradox: the inquirer (or creative agent) does not know what to look for if it is not there already, but might have sufficient information to formulate a question – or requirement – against which candidate solutions might be matched. The requirement may indeed need to be informed by considerable prior knowledge: Kekulé recognised the applicability of his daydream about snakes and closed curves to the molecular structure of benzene against the background of much previous work on organic compounds.

> Even Plato's original formulation of the Meno problem suggests (correctly, in my view) that explicit inquiry is a kind of *search* and that the Meno problem is basically a problem of *recognition* (or construction),[…] namely, how to recognize and select the (or a) correct answer, or at least a somewhat promising answer, out of a noisy background of zillions of potential alternatives.[…] Plato's text also raises the crucial issue of the relevance of old knowledge to the search for new knowledge. (Nickles, 2003, p.64, footnotes omitted).

This passage summarises the recognition problem. But, Plato misses out a step: the formulation of a requirement is what makes prior knowledge both useful and relevant. And the understanding of that requirement is what makes the recognition achievable. Kekulé would not have been able to make anything of his daydream if he had not already known a great deal about organic molecular structures and the puzzle presented by the composition of benzene.

In human creativity the recognition, and the basis of any known requirement, depends upon the presence of a vacant cultural niche – the ground of value – just as in natural selection a genetic variant is successful because there is a vacant ecological niche in the physical environment, which it can occupy. Recognition in turn follows a search (if no search were needed, the recognised output would not be new). This requirement, search, recognition sequence will be central to the construction of a creative process model, in chapter six.

2.8.1 Proposed modifications to the evolution by natural selection analogy

Nickles (2003, p.55) refers to the 'grand generalization of Darwinian evolution', as 'the "BV+SR model of innovation" or "Campbell's thesis"'. Donald T. Campbell, in his 1960 paper (entitled: *Blind Variation and Selective Retention in Creative Thought* …), acknowledges that others[26] had previously taken natural selection as an analogy for trial and error (1960, p.381). There is thus an extensive history to the tradition of taking evolution by natural selection as an analogy for human creativity.

Evolution by natural selection appears to show reasonable promise as a candidate for adoption as an analogy for a creative process but, from the above, it looks as if it may benefit from some adjustment. The following steps have so far been put forward as constitutive of such a modified analogue process:

1. Prior knowledge leads to the formulation of a requirement (not necessarily held in a thoroughly explicit, conscious mode) which could only be satisfied by currently unknown creative output.

2. The novel idea is located as the result of a search (again not necessarily consciously planned or executed), perhaps involving a component of BV+SR.

[26] Campbell (1960, p381) cites as examples Ashby (1952), Baldwin (1900), and Pringle (1951). He also traces the term "trial and error", applied to thinking, back to Bain in 1855 (1960, p385).

3. Recognition of a potential creative solution results from matching the output of the search with the known requirement.

At this stage, nothing has yet been encountered that would bring into question the overall viability of this sequence as a process. The most problematic area seems to be the sheer vastness of the universe of possibility, and of the possibility space attending a BV+SR search. However as Richard Dawkins (1986, pp.43-50) has argued, the cumulative nature of evolution considerably reduces this difficulty.

2.9 Conclusion

This chapter's in depth exploration of the nature and context of creativity has brought out a number of features which will have to be taken into account when planning the construction of a model for a creative process.

Creativity is the fount of novelty and added value in the sphere of culture. Any candidate model will need to be capable of producing output with these properties, and to bring forth the equivalent of original psychological novelty: that which is completely new to the individual, if not to the cosmos. There is some superficial resemblance with evolution, which gives rise to novelty in nature.

Creative products are almost necessarily surprising. This is the other side of the novelty criterion: if it could be anticipated beforehand it would not be new. This needs in some way to be rendered compatible with the search for explanation, which demands traceability to antecedent causes. Therefore as a working assumption, it is taken that creativity does not involve any break in the causal chain, even though new properties may arise that were not present in any prior input state.

A second, not unrelated, assumption is that creative output is sourced from elements of pre-existing mental content, whether internal stored memory or representations of current sensory impressions – in other words that ideational material does not appear *ex nihilo*.

A creative act can be viewed as rendering actual what was previously only possible (and the creative product being brought into existence must be embodied in an empirically possible form, even if it expresses or symbolises higher degrees of possibility). The model will need to overcome the problem posed by the relation between the (effectively infinite) scope of the possible and the need for the isolation and actualisation of potential for added value.

A way will need to be found to circumvent the *Meno* paradox and the problem of search and recognition. It has been suggested that the formulation of a requirement, a facility not available to organic evolution, may hold the key.

These matters can now be taken forward to chapter five, for further consideration in the assessment of the feasibility of constructing the process model.

Chapter Three: *Modularity*

3.1 Introduction

I have completed the initial survey of the overall field of creativity, and also explored relevant aspects of its nature in more detail. In this chapter, I turn to discuss modularity, which is one of the two principles that I intend to use to formulate the model for a candidate creative process. I investigate modularity initially in its original, Fodorian, manifestation, and subsequently in its extended, massive modularity, form. It is in this latter version that it might offer the potential to serve as a platform against which to set the model. I will leave consideration of the second principle, the proposed use of an analogy with evolution by natural selection, to be covered in chapter five.

My overall aim here is to develop a plausible creative process model: modularity might hold the potential to provide an adequate structural framework against which such a process could operate. The successful construction of a viable model on the basis of a modular framework would demonstrate how creative novelty might feasibly be originated, even if the model did not correspond to the way in which human creativity works. Human creativity might operate against some other kind of architecture. It might even conceivably need no mental structure at all (though in that case it is difficult to see how this situation would be compatible with any type of *process* (how could there be a series of steps?).

My goal for this chapter is to understand the case for mental modularity, and to inquire whether it is capable of supporting a creative process. Therefore I am less concerned here with critical analysis of the theory *per se*, except insofar as to establish that there is a credible case in its favour (the human mind being more or less the only known vehicle for creativity). The object is thus to envision what a modular format might look

like, and then from there (in chapters five and six) to go on to explore how such a structure might facilitate a creative process.

I begin by examining Jerry Fodor's version of modularity, which he restricts to peripheral mental input and output systems. Fodor's view of modularity is positioned within the Computational Theory of Mind, the notion that the mind operates along the same lines as a digital computer. This makes it particularly suitable given that the current aim is to design a model for a creative *process*:

> Computational nativism … is primarily about the nature of mental *processes* (like thinking, for example) … (Fodor, 2000, p.6).

My intention here is to understand and apply the principles behind Fodor's system, rather than to critically assess its foundations. I will therefore concentrate on its function and mode of operation.

According to Fodor, central systems are holistic, rather than modular; I consider why he argues that the nature of the central systems' roles implies that they must be unencapsulated (and therefore not modular). The central systems are the only credible locus for creative processing in Fodor's architecture and therefore the case for a modular creative mental process turns on whether reasonable grounds can be established for a modular structure to be extended to at least some central systems in addition to Fodor's periphery.

I therefore move on to examine the alternative theory arguing for more widespread (massive) modularity. I endeavour to show that, with only limited modifications, a quasi-Fodorian module might be enabled to access external data, as is likely to be essential for creativity. The modelling exercise will not need to presuppose that *all* of the mind is modular, only that a modular framework is available where necessary to support a creative process.

3.1.1 Objective: the potential benefits of a modular framework

In sum, if creativity is to be understood naturalistically[27], rather than as a providential or inspirational phenomenon, then some kind of process has to be seen as taking place to generate a creative thought or product, and it is difficult to see how any processing can occur in the complete absence of structure, whether or not this might be innate, or partially so. Relatively little is known as yet either about mental architecture or about processing in the physical neural systems of the brain[28]. The modularity thesis does however propose a structure for at least some types of mental processing, and so this can be taken and investigated to see if it is adequate to provide a framework for creativity.

The purpose of this chapter is therefore to outline and probe the nature of modular structure, so that it can subsequently be assessed for potential as just such a framework. It examines both the modularity and the massive modularity theses, with the minimal objective sought being a candidate mental architecture that might plausibly offer appropriate support for creative processing. Whether or not modules correspond to actual structures to which the mind might be expected to conform is a central question for modularity theorists, but of lesser significance for current purposes. If a coherent framework can be identified which is adequate to support processing capable of generating creative output, then that in itself is a naturalisation of creativity, even if it is not clear that human creativity corresponds exactly, or indeed at all, with the adopted model.

[27] Here, *naturalism* is taken to be the notion that all things are the outcome of natural causes (so that supernatural intervention does not happen). *Materialism* is taken to assume that all things can be explained as the result of material causes. *Physicalism* is taken to hold that there is no reality other than the physical world.

[28] I will not be concerned here with the relationship of the mind to the brain, or how mental modules might or might not relate to physical neural structures, as these issues are not directly relevant to the question as to whether a plausible creative process model can be constructed, given a modular framework.

3.2 Modularity according to Fodor

On a general view, the modularity thesis asserts that the mind incorporates a number of relatively independent subsystems, operating effectively in isolation from each other.

The current debate on modularity originated with Fodor's *The Modularity of Mind*, which he published in 1983. Fodor proposed that the mind is made up of three different types of component: central systems; input and output modules; and also transducers (which convert detected external physical conditions – of light, sound, or pressure, for example – into neuronal impulses).

This section focuses on the central systems, from the viewpoint of their significance for the understanding of creativity (Fodor's input and output modules being unlikely locations for any creative process). It must be said from the start that, according to Fodor himself, the central systems are nonmodular, so the section begins with a brief summary of Fodor's view of modularity, before proceeding to examine his reasons for thinking that the central systems are not themselves modular.

3.2.1 Fodorian mental modules

Fodor identifies nine properties as being characteristic of mental modules, of which the key pair for current purposes are: domain specificity (relating to the range of mental content accepted), and encapsulation (relating to the range of information taken into account in processing the input). Fodorian modules are domain specific in that they only accept input relating to a single type of information (such as visual or auditory sense data). They are informationally encapsulated in that the processing within a module only has access to data held internally to that module. It is the absence of these two properties on which Fodor bases his case for the existence of nonmodular central systems.

It is only the input (the five senses, with the addition of language) and output functions of the mind that Fodor claims are processed by domain-specific, encapsulated modules

– independent, self-contained subsystems each dedicated to their own particular function. They are domain-specific in that they are specialised, each being concerned with one particular area of incoming or outgoing data, and encapsulated in that for example input systems have their own task-specific information base, and have no access to any of the data held elsewhere within the overall system and organism of which they form part.

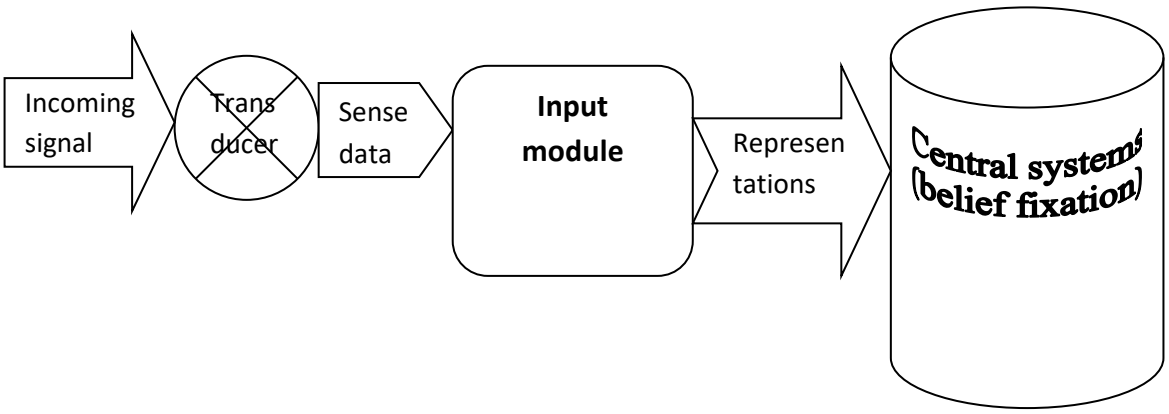
Encapsulation of input systems is necessary, Fodor argues, for veridical perception (1983, p.102). If input systems were open to influence from prior belief or desires, this could lead to the formation of false ideas about the exterior world, through a type of 'wishful' (1983, p.103) perception. The distorting effects of such premature evaluation are avoided, Fodor thinks, by bringing in interactions with beliefs and desires only after the initial processing of inputs has been completed.

3.2.2 Fodor's central systems

Fodor acknowledges that the output from the perceptual and language modules will ultimately require to be assessed in the light of previously established beliefs about the outside world, in order to inform decisions as to how best to respond to the perceived external situation. This means that 'relatively nondenominational' (1983, p.103) (general in scope, as opposed to domain-specific) processes must be brought to bear at some point; Fodor refers to them as: '"the fixation of perceptual belief."' (1983, p.102). He postulates unencapsulated central systems to carry out these tasks.

I have sketched the general pathway for perception and language input in the following diagram.

Fig. 3.1: Fodorian input module



Only one of Fodor's six input modules is shown here. The transducer turns the external physical signal (carried by light or sound waves, for example) into neuronal data, in which form it can be accepted by the input module for processing. Retinal cells would perform the function of the transducer in the visual pathway. The module compiles and synthesises the data into representations of the observed external world, for onward transmission to Fodor's (holistic) central systems, as described below.

Fodor considers the central systems to be differently organised and structured from those input and output systems which he has deemed to be modular. This reflects their distinction in function: input systems are restricted to the computation of representations of what is detected by the senses for onward transmission via their interfaces with the central systems, which Fodor views as being concerned with mediating the fixation of belief.

Each input module takes in sense data and thence constructs and outputs representations. The representations are based solely upon the sensory input, plus

that subset of the organism's information holdings which is available to the module concerned. But, once received into the central systems, the representations then need to be subjected to assessment in the light of other relevant information, leading to the formulation of the best available hypotheses as to the state of the external world. Because of the broader scope, Fodor questions whether this belief fixation process can be domain-specific; it seems that the central systems must incorporate some mechanisms operating across domain boundaries.

Fodor proceeds to claim that central systems are unencapsulated, and therefore cannot be modular. In the absence of direct evidence as to the nonmodularity of central systems, he argues from analogy, taking the initial assumption that belief fixation, by 'nondemonstrative inference', is 'the typical function of central systems' (1983, p.104). Here, nondemonstrative means that the inferences are not arrived at by standard deductive reasoning from premises to conclusions.

Central systems take the representations from input systems together with stored information from memory, and compute '"best hypotheses" about what the world is like' (1983, p.104). All of this generally takes place below the level of awareness, and the processing involved remains obscure. But Fodor intends to use 'empirical inference in science'[29] (1983, p.104) as an example of an '*explicit* [process] of nondemonstrative inference' (1983, p.104), and thus an external parallel on which can be founded his theorising about the central systems.

He begins by postulating 'that scientific confirmation – the nondemonstrative fixation of belief in science – is typically unencapsulated' (1983, p.104), and has the properties of being both '*isotropic'*, and '*Quineian*' (1983, p.105), where:

> ***Isotropic*** means that the confirmation of a scientific hypothesis may involve drawing upon information from any sector of the body of currently accepted

[29] I.e. the process of induction whereby general laws are inferred from observation of particular occurrences, in the way that Hooke's law might be inferred from a series of observations of the length of a spring when different weights are suspended from it.

> scientific theory, so that any of the knowledge to which the scientist has access might turn out to be relevant; and
>
> ***Quineian*** means that hypothesis confirmation is also tested against the background of properties of the body of scientific theory as a whole[30] (see also Cain, 2002, pp.194-195).

Because of the global scope of the range of possibly relevant information, isotropic systems cannot be other than unencapsulated. Also, given that scientific confirmation is Quineian, selections may be made from among competing theories according to criteria of simplicity, for example, which can only be measured against the totality of scientific understanding or belief (1983, p.108).

Scientific confirmation thus seems to offer a case of an isotropic, Quineian, and holistic process, and there are grounds for suggesting that it can be compared with belief fixation within the central systems. Fodor does however acknowledge that it is far from universally accepted that scientific confirmation is isotropic and Quineian, and therefore declares this as a working assumption.

Fodor summarises his line of reasoning in the following words:

> The argument is that the central processes which mediate the fixation of belief are typically processes of rational nondemonstrative inference and that, since processes of rational nondemonstrative inference are Quineian and isotropic, so too are central processes. (1983, p.110).

Fodor then goes on to claim that it follows from that assumption that central processes cannot be modular, since they are isotropic and Quineian, and such systems cannot function effectively if they are constrained by encapsulation.

This allows for a clear distinction to be drawn between (holistic) central systems and (modular) input systems. Both types of subsystem operate by means of hypothesis

[30] See the entry for 'Duhem thesis' in Blackburn (2008, p.106): 'Sometimes called the Quine-Duhem […] thesis. The thesis that a single scientific hypothesis cannot be tested in isolation, since other, auxiliary hypotheses will always be needed to draw empirical consequences from it …'

construction and confirmation, but whereas input modules only have reference to a restricted data set for confirmation, the fixation of belief by central systems needs to be free to access any of the information held by the system, and hence isotropic and nonmodular. Quineianness has similar implications because Quineian confirmation involves comparison across the whole belief system.

Thus having set up the analogy between psychological belief formulation and scientific confirmation, Fodor provides further material by way of validation. He considers the kinds of obstacle that lie in the way of any attempt to formulate theories of central processes, and shows how they are similar to those that would be anticipated if the processes were Quineian and isotropic, and hence unencapsulated and nonmodular.

Fodor cites the “frame problem” (1983, p.112), from artificial intelligence, as an example; this refers to the difficulty of limiting the set of beliefs that may need to be considered for revision when a new piece of information is acquired. He explains that there is no known way of defining such a frame, which is exactly the sort of problem that would be expected if central processes resembled scientific confirmation in terms of being nonmodular and holistic in scope.

In addition, Fodor proposes that, if input analysis is mediated via modules, and more general thinking is isotropic and Quineian, then this could be expected to be reflected in corresponding sections of the brain, in the form of an organised pattern of neurons for language interpretation, and a much looser structure for areas of the cortex associated with thought. He suggests that this does accord with what is actually observed empirically.

Thus Fodor’s argument for holistic central processing can be summarised as follows. He proposes an analogy between belief fixation in the central systems, and scientific confirmation. He further suggests that scientific confirmation is Quineian and isotropic, and therefore unencapsulated, so that if the analogy holds, central systems will also be unencapsulated, and thus nonmodular. This seems in line with findings from

neuroanatomy, and also with the kinds of problem encountered when investigating similar tasks in an artificial intelligence context.

3.3 Discussion: Fodorian modularity

Fodor's book is particularly important in setting out a case for modular mental elements. A module is dedicated to processing one particular type of information, against the background of tightly circumscribed data access and presumably relatively invariant processing routines. These highly structured, fixed computational units seem unlikely settings for creative thought, which is characterised by flexibility and often wide-ranging scope (in the modular context, almost literally 'thinking outside the box'). The properties of modules are thus the opposite of what would be expected of a system capable of supporting a creative process. Also, the criteria for creativity include that the output must be novel and, as Fodor himself notes, creative thought often involves taking ideas or principles developed in one field and applying them in another[31] (1983, pp.106-107), whereas: 'By definition, encapsulated systems do not reason analogically.' (1983, p.107).

So, it seems that the seat of creativity needs to be sought elsewhere, and in Fodor's terms, 'elsewhere' is within the central systems, which (apart from the transducers) seem to be the only other component in his mental architecture.

However, if Fodor is justified in his thinking, the isotropic and Quineian nature of central processes implies that the central systems are opaque to cognitive science, which in turn suggests that creativity is fundamentally unintelligible, along with all other functions of the central systems, and there would be no utility in investigating its mechanism any further. The *prima facie* implications for the current project, therefore, seem to be that modularity (or at least some comparable form of structure) will have to be shown to

[31] For example, the control principle embodied in the punched cards used in the Jacquard loom was later used in computing (see Burke, 1978, pp.111-113).

have the potential for extension into Fodor's central systems, which constitute the only available candidate location that seems capable of mediating creative thought.

In moving on to look more directly at Fodor's method, it may be appropriate to commence with the type of argument he adopts. Fodor chooses to work from analogy, which is a relatively weak form (compared with deductive reasoning, for example), but it must be accepted that, as he points out, there is a lack of evidence to support any more robust line of approach. He backs up the resultant claims by showing that the underlying neuroanatomy seems to be as would be expected, given that the central systems are holistic and unencapsulated, and also that the types of problem encountered when attempting to construct theoretical mechanisms for the central processes are as might be anticipated if they were nonmodular.

These forms of verification are helpful to Fodor's case, as they are fully compatible with his holistic theory. But this is not to say that they provide unequivocal confirmation; they are in themselves insufficient to do this. There is no guarantee that some alternate architecture might not manifest the same effects. As an example, if the central system were to consist of a large number of relatively small micromodules, and with dispersed interconnections, the overall effects might be very similar in terms of the examples Fodor quotes, although originating from a system structure which would be, at least in principle, intelligible.

The analogy is dependent upon two strong assumptions. The first of these is that belief fixation is a typical task for the central systems; but, although Fodor has belief fixation as typical, he does not suggest that this is their sole function. He himself mentions others, including '"problem-solving" (i.e., nondemonstrative inference in the service of individual fixation of belief)' (1983, p.105), hypothesis formulation (as part of belief fixation) (1983, p.102), and '"analogical reasoning"' (1983, pp.106-107).

Fodor does not explicitly cover other tasks that it might be expected his central systems would need to handle, such as: prediction and forecasting, deductive reasoning,

imagination, desires, or emotions. These may be processed in a similar holistic fashion to belief fixation, or they may not: one example of a central process which could in principle be handled by a domain-specific module is provided by mental arithmetic – such routine computation being carried out by a relatively simple silicon chip in a calculator.

The other key assumption is that scientific confirmation (that is, when a theory is taken to be borne out through subsequent observations being in line with its predictions), which Fodor uses as an analogy for belief fixation, is Quineian. This depends on acceptance of Quine's view, otherwise as Fodor says, the requirement for global data access would not apply:

> ... it was ... one of the "dogmas of empiricism" (Quine, 1953) that there must be *semantic* connections between each theory statement and some data statements. ... The effect of postulating such connections would be to determine a priori that certain data would disconfirm certain hypotheses, *whatever the shape of the rest of one's science might be*. For, of course, if H entails O, the discovery that –O would entail that –H. ... To postulate meaning relations between data statements and theory statements is thus to treat confirmation as a *local* phenomenon ... (1983, pp.108-109)[32].

The assumption is thus well documented, and Fodor does consider the possibility that an encapsulated system could actually be Quineian (with respect to a subset of the overall available information holding) (1983, p.111). He also includes a caveat which appears to further call into question the mutual exclusivity between encapsulation and global scope:

> So it seems clear that isotropic/Quineian systems are ipso facto unencapsulated; and if unencapsulated, then presumably nonmodular. Or rather, since this is all a matter of degree, we had best say that *to the extent that* a system is Quineian and isotropic, it is also nonmodular (1983, p.111).

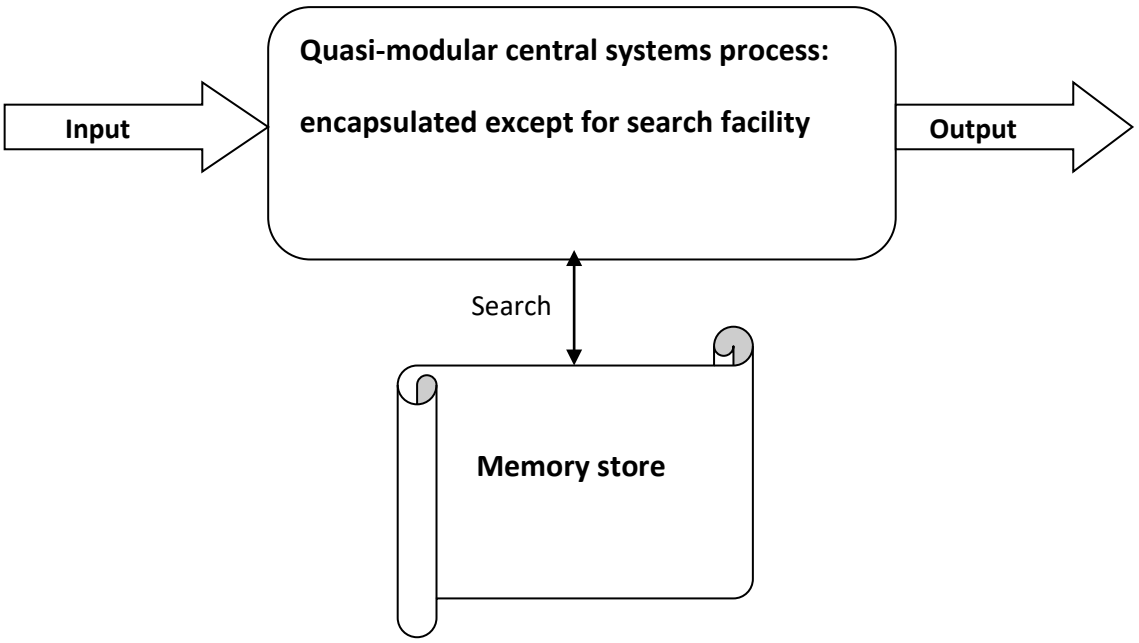
These last points, together with the assumptions, do seem to suggest that there might be some room for manoeuvre, some blurring of the boundary between the modular and

[32] 'Quine (1953)' here refers to the first edition of Quine (1961) 'Two dogmas of empiricism'.

the holistic. Fodor's criteria for modularity, his nine properties of modules, are his own. His characterisation of central systems as holistic depends on these criteria being accepted as inflexible, but is there perhaps some scope for their relaxation, and corresponding elasticity in the modular definition? (This would be a requirement if creative processes were to be accounted for as a function of intelligibly modular, or quasi-modular, structured central systems). Perhaps the central systems can be partially modular, or modular in a limited way, or have some subset of the attributes of modularity?

As an example, Fodor (1983) does not consider whether a module might be domain-specific and encapsulated with respect to processing, but at the same time unencapsulated in the sense that it had a search facility which could access information holdings beyond the module itself, as depicted in my diagram below.

Fig. 3.2: Quasi-modular central systems process (conjectured)



Here, I have allowed for the system to access externally stored data to carry out a search, although the internal processing otherwise only has access to data held within the 'module' itself. The system would not be a module in Fodor's terms, because it is not '"encapsulated" with respect to information that is *not* in its database.' (2000, p.63). However, this arrangement could carry out a search procedure, which will later form an important component of the model (see chapters five and six). I have included the diagram here to illustrate that such an arrangement might be conceivable, and that a process could be global in scope, and yet still be domain-specific, and thus in principle intelligible.

Fodor comments more directly on creative processes (or something closely resembling at least some of them) when discussing analogical reasoning. His analogy for belief fixation is scientific confirmation, but he mentions, in what is almost a digression, that cognitive isotropy is especially evident in the field of scientific discovery, in which he

says there is anecdotal evidence that analogical reasoning is commonly encountered, and provides some examples:

> ... it really does look as though there have been frequent examples in the history of science where the structure of theories in a new subject area has been borrowed from, or at least suggested by, theories *in situ* in some quite different domain: what's known about the flow of water gets borrowed to model the flow of electricity; what's known about the structure of the solar system gets borrowed to model the structure of the atom; what's known about the behaviour of the market gets borrowed to model the process of natural selection, which in turn gets borrowed to model the shaping of operant responses. And so forth. ... (1983, pp.106-107).

He refers to examples in Ortony's *Metaphor and Thought* (Fodor,1983, p.106) (he presumably has in mind the ones in the paper by Boyd (1979, p.359), for example), and to these might be added further cases in Burke (1978, see e.g. pp.81-113, and throughout; also my earlier footnote). Fodor makes an important point about analogical reasoning here, and one which will need to be taken forward when addressing creativity more directly. He continues:

> ... The point about all this is that "analogical reasoning" would seem to be isotropy in the purest form: a process which depends precisely upon the transfer of information among cognitive domains previously assumed to be mutually irrelevant. ... (1983, p.107).

This extract is also reminiscent of Koestler's ideas on 'bisociation' (1964, see e.g. pp.35-36).

A further issue of potential significance for the study of creativity arises when Fodor is discussing the AI frame problem. He quotes from Raphael (1971):

> "(An intelligent robot) ... must be able to *plan* how to carry out a task, and this planning process usually requires keeping "in mind" simultaneously a variety of possible actions and corresponding models of hypothetical worlds that would result from those actions. (1983, p.115).

Fodor goes on to discuss the difficulty of providing for the computing of possible worlds which correspond to the projected outcomes of planned actions:

> This makes it look as though the [frame] problem is primarily (a) how to notate the possible worlds and (b) how to keep track of the *demonstrative* consequences of changing state descriptions. But the deeper problem, surely, is to keep track of the *non*demonstrative consequences. Slightly more precisely, the problem is, given an arbitrary belief world W and a new state description "a is F", what is the appropriate successor belief world W´? ... (1983, p.115).

These references to possible worlds recall their invocation by Nichols and Stich (2003, see e.g. p36) in accounting for pretence behaviour, which was subsequently referred to by Carruthers when discussing creativity in *The Architecture of the Mind* (2006, see e.g. p319, and see also later on in this chapter).

In sum, Fodor's *The Modularity of Mind* is of pivotal significance for the current project, because it introduces the notion of mental modules. This section has necessarily concentrated upon those sectors of the mind which Fodor terms the central systems, as being the most promising site for the processing of creative thought. It has explored his case for arguing that the central systems are not modular, and the implications for the study of creativity which might be associated with holistic central processing.

It is now clear that, while modularity offers a promising framework against which to construct a model for a creative process, in its strict Fodorian form it is incapable of doing so because of the limitations on the range of functions which can be handled. I will therefore move on to consider massive modularity, which proposes more extensive mental modularisation. I have already tentatively suggested one way in which a Fodorian modular structure might be adapted for some aspects of central systems functions. As said, I am concerned more to build an understanding of the way in which modularity might provide a structural framework, rather than to critically examine the theory. I will therefore concentrate particularly on the form of massive modularity put forward by Carruthers, who has a well-developed modular scheme, and who also writes on creativity.

3.4 Massive modularity

The massive modularity thesis does propose an extension of modularity beyond the Fodorian periphery. Evolutionary psychologists such as John Tooby and Leda Cosmides (1992) and Steven Pinker (1997) put forward the theory that the mind comprises many more modules than Fodor envisaged, that there is no central processing system engaging in holistic reasoning, and that the modules are products of evolution and arose out of a gradual process of natural selection in response to the environmental challenges of the Pleistocene. Evolutionary psychology sits particularly easily with a modular theory of mind, as it is the proposed distinct mental subfaculties which it is suggested have evolved, and thus have innate, or hard-wired, aspects. Massive modularity brings with it the possibility of a new approach to creativity, representing it as a process associated with a dedicated module (or modules) having a specific evolutionary function.

3.5 Carruthers and *The Architecture of the Mind*

Unlike Fodor, Carruthers directly addresses creativity. His approach from evolutionary psychology was introduced in chapter one; in his 2006 book *The Architecture of the Mind*, this is linked in with his defence of massively modular mental architecture, which it is hoped will provide a plausible framework against which creative processing can be set. Therefore, the following section considers Carruthers' views as expressed in the book, focusing particularly upon aspects having most bearing on human creativity[33].

Carruthers is aiming to indicate the way in which a massively modular mental architecture could support creativity. This is not the same objective as my current one of attempting to model, and hence understand, a plausible creative process. He does

[33] *The Architecture of the Mind* is not entirely, or specifically, about creativity as such, as indeed its title suggests. It is instead oriented towards three more general aims, as follows (2006, p.xi): (1) to put forward a case 'for a massively modular account of the architecture of the human mind'; (2) to respond to 'a "How possibly" challenge to any such approach'; and (3) to put forward 'at least a sketch' of the links and interactions between the proposed modular components of the mind.

however have much to say about the process of creative thinking. He provides support for his case for massive modularity via an evolutionary psychology line of argument, among other things. His approach to creativity from evolutionary psychology was introduced in chapter one; the following account and discussion concentrate on Carruthers' proposals concerning creative processing, and the plausibility thereof.

Carruthers' modules are non-Fodorian, and are not required to be encapsulated, leaving a 'module' as a relatively self-contained mental subsystem, isolated in terms of its internal processing, and efficient in its calls on data and functions outside the module boundary (see 2006, p.xii). He points out that creative processing is incompatible with encapsulation because creative thought often entails linking otherwise unrelated ideas (2006, p.288, and cf. Fodor's comments on scientific discovery, as above).

Creativity is addressed in Carruthers' Chapter 5, in which he sets out to explain how a massively modular structure could support the characteristically human degree of mental flexibility, focusing 'on the creative and innovative capacities' (2006, p.277, see also p.xiii). He begins by considering the nature of creativity itself; he declines to offer a definition, but describes it as set out in chapter two, above. He has creativity firmly tied to action, and refers to his view as being action first, as opposed to thought first. As in his 2002 paper (see chapter one) he has childhood pretend play as practice for adult creativity, and drawing on the same cognitive resources.

According to Carruthers, creative pretence is founded upon a capacity for originating and rehearsing 'action schemata' (2006, p.xiii, see also p.304). He uses the term 'action schema' to refer to a mental representation of a physical action (this usage presumably deriving ultimately from Kant). Such rehearsals are "globally broadcast" (so that they become accessible to a variety of other mental subsystems (2006, p.84[34]).

[34] Carruthers sources this from Baars (1988).

Carruthers' linking of creativity and action recalls the point made earlier, that creative products have to be empirically possible.

As in the 2002 paper, Carruthers claims (2006, p.281) that pretend play exists to provide practice for adult creativity, the connection between the two being their common cognitive roots. The young of other mammals practise adult behaviour in play, and human pretend play evolved to practise for the imaginative thinking which will subsequently manifest in adult creativity. The shared 'cognitive underpinnings' (2006, p.282) are the ability to formulate a supposition, and then to reason within its confines.

Also as in his earlier paper, Carruthers refers to creativity as having two aspects (2006, p.282): firstly, the facility for the origination of ideas; and secondly, the capacity to appreciate and explore their implications. He cites the "geneplore" model of Finke and others[35], according to which creativity has two steps (2006, p.283): originating and entertaining a new hypothesis, and then investigating and assessing its potential.

Carruthers postulates that creative, suppositional, thought is ultimately rooted in the origination of action schemata. A novel action schema might be applied to generate a physical act, but can also be constrained to produce mental images, which are then globally broadcast. If the schema corresponds to an utterance (in the form of an action) the internalised speech can offer new word strings to the conceptual modules (see 2006, p.287).

Carruthers had previously suggested (in Chapter 2.8, p.145) that, in apes, 'creative generation of action schemata gets driven by associative processes…'. He points out that children's supposition and pretence differ principally in that a broader range of action schemata, which need not be so closely linked to immediate concerns, can be taken to seed the suppositions (see: 2006, p.290).

[35] Carruthers cites: Finke, Ward and Smith (1992); Finke, (1995); and Ward, Smith, and Finke (1999).

On Carruthers' account, the pretence is not 'thought-first'; it is not started by a creative idea, such as: "*The banana is a telephone*"[36]. Instead, he argues that it is more credible that the internal representation (BANANA) of a banana seen by the child can bring about weak activation of TELEPHONE via the likeness of shape. He considers that TELEPHONE would then be able to prime appropriate action schemata (2006, pp292-293).

He has downstream elaborations dependent upon the supposition being "tagged" as such. In this way, the child is able to proceed with the pretence without having to believe that the banana has actually turned into a telephone (see: 2006, p.292).

On Carruthers' view, pretence (and thus also creativity) are associative. He considers that the associative linking could be modelled by postulating 'semantic primitives' as being held alongside, and shared between, concepts (such as BANANA, TELEPHONE). The activation of associated concepts can then proceed via these connections. In earlier sections of the book, Carruthers had favoured computational theories, but he does find a role for associationism in the creative origination of new thinking or behaviour beyond the scope of computation and unrelated to perceptual observation (2006, p.293).

Carruthers gives the activation of action schemata a central role in creativity but he recognises that this presents problems for the imaging of things that are impossible in the actual world (that is, beyond the bounds of empirical possibility), and so could have no corresponding action schema. He gives the example of fantasising about toppling the Eiffel Tower. In this case, he argues that an image of the Eiffel Tower collapsing might be put together in the absence of any prior mental articulation of the event in words as inner speech. Here, he brings in the notion of using an indexical. He proposes that 'an abstract motor schema with the content, *pushing over a tower*, might

[36] Nichols and Stich (2003, p.19) trace a banana-telephone example back to Leslie (1987).

be primed.' The indexical can then be substituted as in 'I PUSH OVER THAT', which could be used in the formation of an image of the toppling tower (2006, pp.302-304).

Carruthers suggests that childhood pretence can generate action schemata for speech, exactly as for other forms of action. As above, the sight of a banana may activate TELEHONE, which is primed by the similar profile. TELEPHONE can then in turn prime relevant action schemata, including schemata for speech that involve saying 'telephone'. This can prompt the formulation of the phrase "Grandma talks on the telephone" (2006, p.305) as TELEPHONE also links associatively with GRANDMA. The assembled phrase is only subsequently comprehended as content.

A pretence episode may also be initiated by inwardly rehearsing a speech action schema unrelated to the child's current context, as inner speech. The child, seeking a topic, searches for 'possible fillers' for speech schemata such as '"I am a… ." [cowboy, nurse, and so on]' (2006, p.307). In this way, Carruthers argues that language provides a facility for introducing novel mental content and meaning, because in this system, the new sentence is put together as words only (devoid of content) in the first instance (2006, pp.310-311). This provides for the creation of novel thoughts that could not have been assembled by starting from the meaning of the created phrase. Carruthers gives as an example of such a thought: '*The number seven is red*' (2006, p.312).

Carruthers compares his approach to creativity and pretence with that of Nichols and Stich (In *Mindreading*, 2003). He takes their account to be "thought-first", in contrast with his own "action-based" version. (He had previously referred to the Nichols and Stich model in the 2002 paper – see chapter one.) Nichols and Stich posit a "possible worlds box" (Carruthers, 2006, p.318, see also their own diagram at 2003, p.36, fig. 2.2, and as reproduced at fig 5.1 in Carruthers, 2006, p.319) to hold representations of possible states of affairs, stemming from pretence, as working memory. A "script elaborator" introduces further suppositions to build upon the opening pretence, but

Carruthers prefers the term “supposition generator”, because it also sets up the original supposition (see also chapter one, above).

Carruthers observes that it is possible to read this arrangement in such a way that it is consistent with his own account, as follows. Carruthers has suppositions as being set up using action/speech schemata. The tagging of content subsequently generated as dependent on the initial supposition could be taken as equivalent to the Nichols and Stich possible worlds box. (However, Carruthers does see differences between the two approaches, for example because the possible worlds box is not connected to motivational systems, but operates with propositional material alone (2006, pp318-321)).

3.6 Discussion: Carruthers and massive modularity

Carruthers bases important sections of his argument on the adaptation of mechanisms which predate human creativity. The evolutionary development of human creative capacity is of minor interest *per se* in terms of the current study, except insofar as it provides insight into the kinds of modular structure which may form the basis against which creative processes are facilitated, or sheds some light on the way in which creativity operates more generally.

Other aspects of Carruthers’ account of creativity do have considerable relevance in the current context. He is aiming to indicate how a modular structure could support flexible and original thinking (2006, p.277). This is more directly relevant to the current study – to demonstrate that modularity can be compatible with creativity and novel thoughts would itself be a significant step forward. Even if the mechanism proposed turns out eventually not to be the one actually used by the mind, feasibility would have been established in principle.

Carruthers' action-based creativity, together with the accompanying views on associationism, and of the significance in this context of language, suggest particularly interesting, and potentially fruitful, directions of approach.

In setting up his action-based hypothesis, Carruthers argues that it utilises pre-existing structures and mechanisms, which is more plausible on evolutionary psychology grounds than bringing in entirely new ones (and in line with the conservatism observed in patterns of adaptation in other species, and for other structures and organs of the body).

Against this, language does represent a completely novel capacity in humans over other animals, and it brings with it a set of syntactic and logical rules for the expression of ideas and concepts. And if this significant new development could be achieved, then perhaps other innovations need not be entirely ruled out.

It is of course doubtful that a new idea which is not expressed can count as an instance of creativity (see Carruthers' distinction between creativity and 'thinking creatively', as mentioned above). Any form of expression must involve action in some form or other, so in that sense it seems plausible that Carruthers' action schemata, or something like them, must be involved at some point in the process. Speech action schemata could conceivably drive the physical movements involved in making an utterance. However, it is more difficult to see how they would have any bearing on the grammatical form of a sentence, or on its content in terms of the meaning intended, or as understood.

Language has more than an action component, it has syntax and conveys meaning, and these abstract elements must also be present in any form of creative expression involving linguistic utterances. It does not on the face of it seem necessary to invoke action schemata to compile the content-independent sentences envisaged by Carruthers. No alternative mechanism immediately suggests itself, but since language is a novel development in humans relative to animals, and since it involves both these

elements, perhaps it does not seem unreasonable that language-specific cognitive systems capable of such manipulation evolved alongside language itself.

Carruthers' version of creativity is specifically not thought-first, but the abstract generation of action schemata he invokes, does seem to bear some resemblance to a thought-first process. The distinction between the manipulation of abstract action schemata, on the one hand, and thinking in words, on the other, seems to be relatively subtle – perhaps the crucial difference lies in the absence of reasoning, such that sentence construction can precede understanding sentence content.

Carruthers is correct in noting that content-independent sentence construction expands the range of sentences which can potentially be put together beyond the bound of the constraint to be empirically possible (which must apply to schemata corresponding to actual physical actions), and represents a potential component of a mechanism for generating novel thought content. Sentence construction independent of meaning is freed from constrictions imposed by any logical or deductive rules for manipulating content, or by relationships between the underlying concepts, or by preconceptions derived from current beliefs or prior experience. It is of course debatable whether any process of deductive reasoning (from premises to conclusions) can result in genuinely novel ideas entirely unaided. The absence of reason is also highly reminiscent of Plato[37].

Carruthers' version of massive modularity involves considerable relaxation of Fodor's criteria for modules, in particular on the degree of encapsulation, but as he points out, creative processes frequently appear to involve relating information taken from more than one domain of thought, so that strictly Fodorian modules seem unlikely to be capable of forming an adequate basis for creativity. This is where Carruthers' view that some creative thinking involves association rather than computation seems particularly promising.

[37] See the discussion of Plato's views in chapter one, above.

Carruthers' description of concepts as being connected via semantic primitives recalls the organisation of data structures in computing (e.g. in relational databases), and the formulation of new ideas through association also seems to bear some resemblance to a process of linking across structured databases.

More generally, associative linking across separate domains of thought recalls Koestler's bisociation process (as noted earlier in this chapter).

As Carruthers himself notes, his principle of the tagging of beliefs as dependent on a supposition could be taken as equivalent to setting up a model of a possible world (including the fictional worlds constructed for story-telling). Also, the manipulation of sentences independent of content allows the construction of sentences referring to thoughts which are impossible, or only logically possible, such as Carruthers' own example (quoted above): *the number seven is red*. This is perhaps suggestive of one kind of process which might be considered to have potential to contribute towards instances of Boden's class of creativity involving transformation of conceptual spaces (see chapter two, above).

Carruthers mentions the speed of transmission of nerve impulses (2006, p.24) as imposing a constraint on the tractability of mental processing. These chemical and ionic signals travel much more slowly than electrical currents in a microchip. It is difficult to see that operating at this kind of speed could allow for the rapid formulation of new ideas, either computationally or associatively. It is perhaps notable in this context that Kekulé's idea occurs to him in a daydream, presumably a less than fully conscious state. Moreover, Kekulé's insight (1) occurs in a dream; (2) appears to involve the association of ideas from separate domains of thought, which might be regarded as being linked by 'semantic primitives' (snakes and molecular chains both have string-like properties). Perhaps there is a case for proposing that slower, more resource–hungry mental processing might take place in such states, when there is a

lesser need for rapid response to environmental pressures (see also the discussion in chapter one on Renaissance views of creativity).

3.7 Carruthers' implied process model.

Carruthers provides boxology diagrams to accompany some parts of the book (see pp.141 & 233, for example), although not for creativity. Some of the points he makes might, however, be reinterpreted as processing steps; examples follow.

His two aspects to creativity, the generation of new ideas, and the development of their significance (2006, p.282), recall not only the criteria of novelty and value but also the variation and selection elements of biological evolution, which can be viewed as successive stages in the natural selection process.

Searching can be involved in Carruthers' system, for example in finding fillers for sentence schemata (although he also alludes to random generation, see 2006, p.307). Searching data holdings will feature in the modelling exercise, in chapter six.

The relations between concepts which allow one (BANANA) to prime others, such as (TELEPHONE) imply links allowing the perceived associations to be stored and accessed. This suggests that the data is effectively structured according to such links. The priming of concepts across such associative links resembles a query of the structured data, seeking such links. The 'structure' formed by the associative links might be assumed to make the search process relatively efficient, or tractable, because the links are already in place, prior to the search. The effect is, perhaps, comparable on a smaller scale with that of indexing the *Universal Library* (see chapter two).

Carruthers' arrangements for the generation of suppositions and the tagging of dependent beliefs allows for the mental exploration of the suppositions by way of possible world simulations. This point will be taken up again in connection with the modelling exercise, in chapter six.

3.8 Summary of main points

I have to assume massive modularity, or I would be left without a structural framework against which to construct any model creative process; in other words, I am unable to make any further progress towards my objective in the absence of such a framework. Therefore, my aim for this chapter has been primarily to establish a good understanding of the principles underlying modularity theories, and only secondarily to apply critical assessment. From the outcome of this exploration, I conclude that modularity theories are sufficiently coherent and well developed to permit the formation of an appropriate structural framework for modelling. Additionally, Carruthers' has demonstrated that a massively modular approach can be used to support a theory of creative thought and action.

I do not need to prove, or even to assume, that modular theories correspond to real mental structures, for current purposes. The successful construction of a viable creative process model would constitute an advance, whether or not it could be shown to duplicate, or to operate along the same lines as, any variety of actual human creative process.

For now, the most significant findings of this analysis of Fodor's and Carruthers' versions of modularity are that:

1. A case can be made for modular mental architecture, at least for some functionality.

2. Also, Carruthers has put forward a plausible and coherent theory of creativity which sits in a (massively) modular context.

3. Carruthers' view of creativity can be taken to suggest steps within a creative process, or at least some elements for such a process. (Further consideration of Carruthers' views on creativity will follow in chapter six.)

These points provide support for the proposed use of a modular framework to construct a model for a creative process. The modelling exercise will not need to presuppose that *all* of the mind is modular, only that sufficient modular elements are available to provide a structural framework against which to set this particular process.

At this point, the Initial set of investigations into creativity, its context, and mental modularity, have been completed. The next chapter carries a distillation of the outcome of the foundation stages of the project. The findings thus far will be drawn together, preparatory to proceeding - in chapter five - to assess the feasibility of modelling a creative process, and also to discuss the modified Darwinian analogy on which the model is to be based.

MIDWAY: LINKING FOUNDATIONS TO MODELLING

Chapter Four: *Formulating a Candidate Creative Process*

In this chapter I will gather together the conclusions from the preceding exploratory investigations to prepare for the modelling exercise to follow. I am aiming to produce a plausible model for a creative process which will contribute to an explanation as to how creative novelty might be generated, in the way that the Darwinian variation, selection, and retention process explains organic evolution and the origin of novel forms of life. It is the origination of novelty, the question as to how creativity can happen at all, that lies at the centre of the mystery and which needs to be explained, and so the modelling exercise will concentrate on that aspect. The origination element will need to be embedded within an overall, total creative process which will plausibly be capable of outputting creative products, not merely new or original ones.

As said, I will first summarise the conclusions thus far. To recap briefly, I arranged the coverage of the first three chapters as follows:

1. I began with an exploratory investigation of the general field of creativity together with its associated questions and problems.

2. I then focused upon the nature and characteristics of creativity itself and of the creative process, in order to determine as closely as possible what needs to be modelled and/or explained.

3. Finally, I considered the potential that mental modularity theories might offer in providing a structural framework against which the model could be put together.

The output of any process mediating the origination of creative novelty would have to satisfy appropriate criteria, in order for that process to be classified as creative. I therefore set out to investigate the principal features and scope of creativity and, in view of its ubiquitous occurrence, I needed to cast my net widely at the start. I designed the earlier part of the study bearing this in mind, so as to inform the more detailed consideration of any candidate process in this and later chapters. I would need to be in a position to inquire as to whether any modelled process would address the element of mystery surrounding the origin of creative novelty, in particular concerning the appearance of new properties apparently *ex nihilo*. I intended to take forward the distilled outcome of these initial explorations to focus the modelling exercise.

I found in the literature no known, or generally accepted, process, no series of logical steps leading from any antecedent state of affairs to a culmination in a creative product, even though creativity occurs in all areas of human activity, holds a key position in the development of the artificial environment, and carries ontological significance. I noted that its investigation incurs cross-discipline risks due to its necessarily broad range. I saw that I would need to establish and distinguish between core and peripheral issues, and decide the line to be taken on them for the modelling exercise.

In the first chapter therefore, I sought to understand the overall setting, highlighting the points that would merit more detailed consideration. The coverage was necessarily exploratory at a high level, taking a preliminary overview of the general field of creativity. This was needed because of the conceptual complexities introduced by the widespread occurrence, and general obscurity, of creative processes and their output, which called for an extensive, as well as intensive, search of the relevant literature.

I thought the scale of the topic suggested that it might be most efficiently approached by taking the most general picture as the starting point. I examined the backdrop in

nature and culture, observing that creative novelty is a subset of all novelty; and a creative process is a channel by which novelty is mediated. Creativity generates new properties and added value, and in this it has some commonality with biological evolution, a natural process known to result in new and adaptive properties (fitness being close to - or at least not dissimilar to - value). But humans are designers, and have new ideas, so where do these innovations come from, given that creative novelty effectively excludes accessibility in advance and new properties may appear to arise *ex nihilo*? Also, creativity is often regarded as a distinctively human capability, although there is no *prima facie* reason to think it could not be implemented in an artificial system.

Darwinism allows for the origination of novelty, as does creativity. The variation and selection components of Darwinian evolution recall, and align reasonably well with, the novelty and value criteria for creative products. Evolution therefore seemed a promising candidate analogy for creativity, even though in nature, new properties are generated by natural law and without the need for a designer. In this context I compared the drivers of natural selection with those of creativity. I noted that the selective factors present in the environment influence the direction of evolution; so that environmental challenge resolves into a function in the generation of novelty through Darwinian evolution (albeit that the interaction between living things and the natural environment can lead to suffering for the individuals thus exposed). I considered whether creativity might be taken as a response to more general natural, or cultural, environmental (or self-imposed) challenge, as in Kronfeldner's definition (see chapter two), the overall trial presented by the human environment. The degree of success of a creative product as a response to the human environment is likely to become reflected in the perceived added value generated by the corresponding creative process, just as the degree of fitness of a genetic variant relative to its natural environment is likely to be mirrored in the extent to which the variant is successful, other things being equal.

I can restate these points more succinctly by saying that:

1. evolutionary novelty arises partly ***as a result of*** challenges presented by the natural environment (variance factors will include a component that is disconnected from selective environmental factors, see later on); whereas

2. creative novelty arises partly ***as a response to*** the natural and cultural environment

I have used more active phrasing for the latter to reflect the design element, which I will need to take into account when constructing the model creative process. Kronfeldner includes self-generated challenges; perhaps these constitute a second group unconnected with the environment, or they could be regarded as being presented by the internal mental or psychological environment, in which case perhaps they are ultimately developed from observation of the external environment. But whichever may apply, the value criterion as applied to a creative product can only refer to the (external) cultural and/or natural environment.

An analogue of the key role of environmental challenge in natural selection will be seen to feature, although by a modified route, in the modelled creative process, as described later on. The influence of environmental testing can also be discerned in the context-sensitivity of the value criterion for creative products. The presence of a need to be fulfilled in the cultural or natural environment invites a creative response.

In this respect, I regard creativity as being closer to artificial (as opposed to natural) selection. In artificial selection, environmental challenge is replaced by a (human) requirement for particular characteristics or properties to fulfil a desired purpose or goal (wheat with a shorter stem, or cattle with a higher milk yield, for example). The properties to be sought can be known, whereas in natural selection the variation is blind. This point will be taken up later on.

I also explored another link with Darwinism, noting arguments in evolutionary psychology in support of explanations for the development of human mental functions (including creativity). I take them as being a separate, although given the context not unrelated, application of the Darwinian principle. But although this approach may provide clues to the nature of creative process (particularly as the human context is effectively the only one we have to go on) I saw it as being of particular relevance to the development of specifically human creativity, rather than for understanding the creative process itself (independently of platform). Evolutionary psychology is, however, of importance in providing an argument for modularity, which will have a place in the modelling exercise.

Creativity and Darwinian evolution both generate novelty, and are thus associated with a more general problem of origination, the question as to how new things are brought into existence. In the case of creativity, this incorporates the more specific point as to how (by what process) ideas and artefacts previously unknown and non-existent, come to be. Here (and as throughout) I am taking 'creativity' in an inclusive sense as a way of thinking (and thus a mental process), not restricted to significant advances or innovations in art, science or technology.

But creative process is resistant to direct inquiry. Anecdotal and introspective reports are unreliable and not very informative in terms of understanding the process and thus of limited value in this respect. Eureka moments are not under the command of the will; they can occur when the individual is not engaged on the issue concerned, as an insight whose immediate precursors are not readily discernable (if at all) or can seem to come from nowhere or nothing.

The output of a creative process, on the other hand, is potentially publicly observable (and so available for inspection and assessment, whereas the process itself is not) and must also be empirically possible. Novelty and added value are the most basic criteria which must be satisfied for any process to be classified as creative, and they both refer

to output. Early on, therefore, I took an initial overview of the novelty and value criteria, and their context-sensitivity. I noted that a creative new idea has to be original, and not copied, and that nonsense, and novelty in the absence of value, are insufficient for creativity.

In addition to any process and product, human creativity can only happen through an agent. Also, and supplementary to the process, product, and agent components, I brought in the notion of the *source*, which I intend to take forward into the modelling exercise. I argued that some form of source was implied given no metaphysical or *ex nihilo* input. I further conjectured that in creativity the source could only be taken from pre-existent ideational material, in which I included incoming sensory information.

The source input cannot, of course, on its own explain the appearance of new properties, only the potential for new combinations with which those properties might be associated (as is equally the case for evolution by natural selection). I also noted that the novelty criterion demanded that the output, or product, of a creative process be inaccessible in advance (otherwise it would not be new). This requirement for spontaneity poses a key question for the understanding of creative process: how can the unknown be the target of any search? Or, how can novelty possibly be originated? This in turn links to the *Meno* paradox, which asks how a solution, unknowable beforehand, can be sought when there is no indication where to find it. Plato puts it down to before-life memory (and in this context I drew attention to the superficial similarity between memory recall and creative insight, in terms of spontaneity). This also relates to the apparent obscurity of any creative process. I noted the paradox that, given no *ex nihilo* input, a creative process (and its product) should be explainable in terms of causes, but then if those prior causes were known in advance, the effect might be anticipated in them (in the way that a syllogistic conclusion might be viewed as being anticipated in its premises). This seems to rule out any straightforward deductive procedure as a candidate for modelling a creative process.

Darwinian evolution resembles creativity in that novelty arising through its action may have been absent in any prior state, but it generates the novelty by the action of random or semi-random variation. In the modelling exercise, I will endeavour to show that a creative process differs from Darwinian evolution in this respect, due to the human ability to anticipate the required properties (although not the creative product itself).

The creative product not only has to be new, it must also add value to the cultural or artificial environment. Here, there might be another similarity with Darwinian evolution, in that value is close to fitness, as both concepts are relative with respect to their context, or environment. The current study focuses on the origination of novelty, rather than the assessment of value, but any modelled creative process will also need to incorporate this aspect, in order to meet even the basic criteria for creative products. I noted that Nichols and Stich have proposed a possible worlds box, allowing for exploring and assessing the potential of a pretence. I also observed that: *to believe that P* builds a model of the actual world, whereas *to suppose that P* builds a model of the possible, which might then be used for the virtual testing and selection of creatively novel ideas (in effect, assessing their potential to add value), a facility not available to organic evolution. I therefore propose to base the selective aspect of the model on this type of arrangement; in this I will also be following Carruthers, who took a similar line.

As said above, I saw potential in regarding a creative process as being centred around a search for required properties, as opposed to a required product. This leads to a further question as to what the target of such a search might look like. I had already proposed that the source input for a creative process must consist of pre-existing ideational material. I noted also the diversity of the properties of words, as being connected by sounds and symbols as well as meaning: I commented that this hints at some kind of structured data storage, as words seem to be retrievable from memory according to these relationships. Such a structure would greatly facilitate searching, as

discussed below. Information storage is thus the most obvious connection I have been able to propose between language and creativity. Metaphor and analogy require the linking of concepts across domains (in a similar way to much creative thinking) and structured information would again facilitate searches for such links.

Finally, towards the end of the first chapter, I began to ponder as to what kind of process or method might possibly be expected to lead to a creative outcome. I noted that Wallas's widely accepted model has four stages: Preparation, Incubation, Illumination and Verification. This was to be given further consideration later on.

Chapter one necessarily took a preliminary excursion through those subject areas neighbouring, and most relevant to, creativity and creative process. Having thus set out the groundwork, in chapter two I focused in on the nature and characteristics of creativity, and its role in the origination of cultural novelty. I was going to need a deeper understanding in order to establish as far as possible the precise specification of what it was that would have to be modelled, and what a model creative process would need to be capable of. I would also have to investigate current thinking about the process itself.

Also in the second chapter, I inquired into and discussed the identifiable varieties of creativity and novelty. I covered relevant aspects of the relation between possible and actual, as well as the ontological problem of novelty and its relation to antecedent conditions. I considered possible sources for new ideas and the difficulties that would be associated with any theory involving random combinations of pre-existing material. I investigated the association between the *Meno* paradox and problems of searching and recognition.

I also looked more directly into the creative process. I took the essential components of any creative process (from earlier on) to be: the (assumed) source; the agent; the process itself; and the outcome or product. The core processing being largely unknown, my starting position was to set out to establish the core and optional

attributes that would need to be replicated in a model, and thence to begin to determine reasonable expectations for modelling. I saw that, when attempting to identify aspects needing to be modelled or explained, the most obvious target area would be the creative output (as it would have to be observable). I considered definitions and descriptions relevant to this area with the aim of crystallising criteria likely to prove sufficiently well-defined for modelling, while not being too far removed from creativity and creative process as normally understood.

I observed that the novelty and value criteria invite qualification and refinement. This applies particularly to novelty in that novelty has to be original, while retaining the requirement for added value (thus excluding nonsense text, for example). I saw that novelty and value both depend on context: novelty relative to what has been brought into existence previously (or is *known* so to have been), and value relative to the cultural environment (although this is of lesser importance for the current study, with its emphasis on the origination of novelty). The two basic criteria are thus not fully sufficient for creativity, but I noted, and agreed with, their more or less universal acceptance as necessary. Novelty and value go a long way towards characterising creativity, even though novelty, in particular, needs refinement. I therefore took them to be serviceable as initial yardsticks for modelling, so that I considered they may reasonably be taken as diagnostic of a candidate creative process, in spite of their respective deficiencies.

In continuing my consideration of suggested definitions, I noted in particular that Kronfeldner links creativity to environmental challenge. This would be of some significance for the subsequent development of the model, in relation to the search for required properties.

Next, I considered the varieties of creativity, which are largely dependent on the type of novelty generated. The current project focuses on the origination of novelty and so I therefore examined the varieties of novelty in addition to those of creativity.

I accepted the anthropological, psychological, and historical forms of creativity (as previously identified by others), which nest into series that are asymmetric with respect to the individual or universal viewpoints. I observed that all these categories require the production of the corresponding type of novelty. I noted that the psychological variety is particularly important for understanding the creative process, because the process can be taken to be the same whether or not the novelty is historic (and thus ontologically significant) provided that it is psychologically original, so that output is generated that has not been copied or learnt. Therefore it is the origination of psychological novelty that has to be explained and will thus need to be modelled.

I then inquired into relevant points on the ontological background to creativity. I noted that creative novelty applies to ideas, often as expressed in artefacts, and thus to culture as opposed to nature. A creative process adds novelty and value to culture, much as evolution adds novelty and adaptiveness to nature. The means of preservation and transmission of new variants differs between the two processes: in culture this is achieved through recording, memory, communications, and artefacts. In nature, evolutionary variants are transmitted genetically.

I noted that Mill's Nature (the inclusive version) encompasses possibility as well as actuality, and that creativity can be regarded as rendering the possible actual, although the possible element must also have been previously unknown. Inclusive Nature is equivalent to what I have termed the universe of possibility, all that is possible under the (actual) laws of nature. Counterfactual and higher degrees of possibility can be communicated but not embodied in creative output. This therefore delineates the scope of potential creative output: it must be empirically possible, originally new, and add value to the human environment or culture.

I observed that metaphysical creativity (whether or not external intervention is invoked) would generate output *ex nihilo*. In such cases, any novelty produced would be entirely disconnected from antecedent conditions. The novelty would not beforehand have

been within the universe of possibility and therefore its generation would be amenable to neither explanation nor modelling. Conversely, if the creative generation of novelty can be modelled, then creative action in the absence of any *ex nihilo* input is at least conceivable. The model may also need to allow for serendipitous (as opposed to *ex nihilo*) input from sources external to the creative individual. The modelling exercise will have to be restricted to higher level processing but the model will need to be fully naturalistic, nonetheless.

In terms of causal independence, I observed that creativity is not independent of prior knowledge and information (in the way that Kekulé could not have found a molecular structure for benzene without prior experience in organic chemistry): it is the exact nature of the relationship that is in question. This is not incompatible with my own view that novel ideational material must arise from the rearrangement of pre-existing sources, rather than be created from scratch. The surprise element required by Boden might, in some cases at least, derive from novel properties which may arise as a result of creative action and which could not necessarily be anticipated from prior knowledge alone (as in the example of a mechanical wristwatch, which tells the time whereas alternative combinations of its constituent parts do not).

I have treated Boden's exploratory and transformational varieties of creativity (which are based on categories of surprisingness) as bringing in a separate principle. They refer to the manipulation of conceptual space; this is a different way of exploring possibility space, and an alternative approach to creativity (on the assumption that they cannot be reconciled with the rearrangement of pre-existing source elements). I suggested, however, that Kekulé's discovery might be viewed as just such a recombination.

I moved on to consider the potential input sources for a creative process, whether internal, external, or located nowhere (that is, *ex nihilo*). The naturalistic view denies any *ex nihilo* origin, so that this source can be eliminated for the purpose of the

modelling exercise. I set up the assumption that all possibilities are contained in or encompassed by the (conjectured) universe of possibility, so that creativity is constrained by its boundary.

In evolution by natural selection, the input sources are pre-existing physical material. I have assumed that the corresponding sources for creative process, if they exist, can only be pre-existing ideational material, so that both types of source give rise to novelty, via the respective process. In evolution by natural selection, new arrangements of physical antecedents give rise to new properties, which I here regard as being latent in the universe of possibility until they are brought into actuality through evolution. I view new properties generated in creative action as transiting from the possible to the actual in the same way.

I argued that, if historical novelty resulting from creative action is produced within the laws of nature, then the source input must be within the cosmos, whether that source is situated within or outside the individual agent. External sources might include: artefacts, recorded or communicated material, or serendipitous discovery. All of these would of course be in the form of internally held data or representations at the point of access by any (presumed) creative mental process.

I then began to examine the notion of a creative process as a series of steps leading from source input to creative output, ideally on similar lines to the elegant, simple, and straightforward ones leading to the origination of successful new variants in evolution by natural selection. Such a series would explain how creative products might plausibly be generated naturalistically, in the absence of any *ex nihilo* element.

I explored the relation between the universe of possibility and the actual world. The vast size of the universe of possibility (and also of an individual's internal data store of prior knowledge and beliefs) renders random combinations unhelpful and hopelessly inefficient (the *Universal Library* problem). In natural selection, this is solved by the cumulative nature of evolutionary change, although the scope for subsequent novelty is

thereby restricted. I noted the lesson of the *Universal Library*, that accessibility to all possible combinations of the available characters is effectively equivalent to having no information at all. In the same way, it is vanishingly improbable that the equivalent of creative novelty could be originated by the random generation of text; although this method was perfectly capable of producing new combinations, it would fail on the value criterion.

I then returned to the *Meno* paradox and the problem of recognising a previously unknown solution. I noted that the question does not arise in the case of organic evolution, because new variants are tested directly against the prevailing environmental conditions. This is a key difference between the two ways of originating novelty; such differences will later be taken forward to inform the modelling exercise. I pointed out that another way of viewing the distinction is to say that in evolution there is no (mental) recognition but adaptive variants succeed, or are (analogue) 'recognised' as fitting an environmental niche.

I also noted that by taking random combinations it is theoretically possible to get out more than was put in: randomly selected text could potentially turn out to be meaningful. In evolution, random mutations form an input resembling the role of serendipitous discovery in creativity, which also has a random element.

Newly emergent properties are latent possibilities until the 'right' combinations are found: they are outside the search space, which can only cover the range of combinations of the source constituents, not any emergent properties absent from the sources. Novel, emergent properties lie beyond the set of output combinations alone. For example, in the case of Kekulé, neither snakes nor open chains of carbon atoms possess the property of being an appropriate structure for benzene. This new property is therefore located beyond the scope of the source. The structure lies in a different dimension of possibility space (aromatic compounds) and its discovery opens the dimension up for exploration in a manner exactly as Boden describes.

In addition to this element of discovery, it is necessary to recognise the value of novel creative output. For that we need a known requirement (which is *not* available to evolution by natural selection) against which to compare the product. Recognition follows a *search*, which must to an extent be blind because it is not known what is sought (or it would not be new, as above). But the blindness applies to the embodiment and not to the properties, which *can* be known beforehand. The requirement therefore specifies the known properties. This is the step missing from Plato's *Meno*: the formulation of the requirement allows the benefit of relevant prior information to be brought to bear, and makes recognition achievable.

It is therefore very much a case of what is searched *for*. The solution, the creative output, is not known, so it is no good looking for it: if it is to be new, it cannot be found in the sources. The target, or focus, of the search must therefore be displaced onto the requirement for a set of properties, which can be based on prior information and therefore known in advance.

I will now summarise the findings from the second chapter. Evolution by natural selection appears to show reasonable promise as a candidate for adoption as an analogy for a creative process. However, a number of points of difference between evolution and creativity have become apparent. These are sufficiently significant as to suggest that a simple, straightforward analogy will not form an adequate basis for a creative process model. Further steps may have to be added to the evolutionary and Wallas schemes. These will need to provide for prior knowledge to be used to inform the compilation of a specification of required properties, which can then be used to conduct a search. The known requirement will allow for recognition, because the output of the search can then be matched against it. The creative process model will also need to be able to assess added value, in order to be capable of generating the equivalent of creative novelty. The vastness of possibility space may present problems, and the process will have to take this into account. Creative products are

almost necessarily surprising: this is the other side of novelty criterion. If they were accessible in advance, they could not be new. But, to be explainable, a creative product still needs to be traceable to antecedent causes and thus sourced from pre-existing mental content (even if that is derived from sensory information concerning the current environment).

In chapter three, I turned my attention to modularity, in order to assess its potential as a structural framework for modelling a creative process. I judged it to be of lesser importance for this purpose that the modular scheme should correspond closely to any actual mental structure. But what I was looking for was the potential to assume a set of more or less independent subsystems, each operating in relative isolation from the others.

I began by investigating Fodor's modular theory. I noted that Fodor had restricted modularity to input and output systems, which were not as such suitable candidate loci for creative processing. Fodorian central systems are non-modular, with holistic processing, but this is no help for modelling. Fodor's domain specificity and encapsulation criteria for modules also presented obstacles, because a creative process may well need to access data external to the module (such as prior source information, see above). I concluded that a creative process could not, in that strict sense, be encapsulated. Massive modularity proposes more extensive modularisation, thus providing some prospect that modifications of the Fodorian criteria might be conceivable.

I noted that Carruthers relates creativity to pretence, and that he incorporates supposition generation in his theory of creativity. I interpreted his view as also moving in the direction of implied structured data, which (as above) I had considered would be useful for efficient searching and might help to offset the *Universal Library* problem. I also noted Carruthers' use of 'tagging' to denote thinking dependent upon

suppositions, and its equivalence to possible world building, primed by current circumstances or prior thoughts .

I registered Carruthers' view of creative thinking as associative (which is not incompatible with mentation being computational in my view). He considers that associative processes will be needed for creativity because it involves relating information from more than one domain of thought, and he is of the opinion that this mechanism would be compatible with the need for tractability. I was also particularly interested by his semantic primitives system which I consider approximates to structured storage of data.

I can now draw together three final sets of conclusions from the investigations in the first three chapters.

1. A case can be made for modularity to be taken as a basis for a structural framework for assembling a model creative process, although an unmodified Fodorian scheme would not be adequate for the purpose.

2. The indications thus far are reasonably favourable towards putting together a plausible account of creativity in a modular context, based on an outline analogy with evolution by natural selection. It has become clear however, that there are significant differences between the two sources of novelty, and so it already seems unlikely that a simple, straightforward analogy will be appropriate. I can now summarise the points of similarity and those of difference as follows.

    a. One thread which has been noticeable throughout, is the way in which the generation of new ideas and the assessment of their significance recall the novelty and value criteria for creative products, and also the variation and selection elements of biological evolution. This suggests that the basic two-stage, variation followed by selection, evolutionary

process can be expected to underlie creative thinking, as some theorists have claimed, and these correspondences will be taken forward to inform the modelling exercise.

b. Both processes can originate novelty and generate new properties not present in any prior state, but creative novelty applies to the human, artificial, or cultural environment, whereas evolutionary novelty applies to the natural one.

c. The input sources for an evolutionary process are pre-existing physical material or organic forms. Those for a creative process can only be pre-existing ideational material (given that there is no *ex nihilo* component).

d. Evolution by natural selection involves random, blind variation, which is subject to selection by direct exposure to the environment. In creativity, prior information is needed to recognise a potential creative idea, which can then be matched against a known requirement for a particular set of properties. The known requirement also allows the new set of properties (although not their embodiment) to be actively sought (as is the case in artificial selection). This is a major departure from the evolutionary base, and the key to the *Meno* paradox, as applied in the current context.

e. The search for required properties will need to cross domain boundaries, and will be greatly facilitated if the prior information to be targeted is structured.

f. I was interested by Nichols and Stich's use of the script elaborator (or supposition generator) and possible worlds box; I intend to adopt a similar arrangement when putting the model process together. I noted

that this is equivalent to creating and reasoning with imaginary scenarios (as in Carruthers).

3. I can also begin to suggest some steps for a potential creative process, based on the above, and as outlined at the end of chapter two. To the basic two-stage evolutionary model needs to be added a facility for searching prior information to formulate a requirement for a (currently unavailable) novel set of properties. This will need to be accompanied by a query function to search across domains for instances of the properties concerned. There then needs to be allowance for the recognition of search outcomes which come close to satisfying the requirement. Finally, there will need to be an arrangement similar to the Carruthers/Nichols and Stich arrangement of a supposer and possible worlds box, to assess the potential of candidate solutions to add value and thus classify as creative.

These points will be taken up during the modelling exercise, for which I will prepare in the chapter to follow. Having commenced with these early findings, I will move on to assess the overall feasibility of proceeding to construct a model along the lines proposed. I will also develop the principles that will govern the theoretical basis for the model.

PART TWO: THE CREATIVE PROCESS

Chapter Five: *Feasibility and the Darwinian Analogy*

5.1 Introduction

This chapter splits into two major divisions. The first concerns the modelling methodology and assessment. The second aims to explore the proposed Darwinian analogy and to subject it to rigorous testing, and also to confirm the conclusions from earlier discussions as to modifications that may be needed to allow for perceived differences between evolutionary and creative processes. The points to be covered include:

1. the objectives for, and appropriate expectations from, constructing a model.
2. the benchmarks that any model of a creative process has to satisfy;
3. natural selection as a process and as an analogue for creativity, and its implications for understanding and modelling a creative process.

Thus, in the current chapter I investigate various factors impacting the feasibility of taking the modular line forward to model a creative process. I intend to use boxological diagramming as a technique for modelling mental process and function, and to take evolution by natural selection as a candidate analogy for creativity (from chapter two, and as further discussed later on). These, together with a modular structure as framework, will form the basis from which to construct a model. Any successful model will have to be compatible with naturalism (as it would be impossible to model intervention) and so would constitute a naturalisation of creativity insofar as creativity is, or could be, mediated by that particular process.

In the earlier sections, I consider some of the more general factors which will have to be taken into account in making an assessment of the viability of model construction, and which may be taken to indicate the degree of success of the outcome of such an exercise.

In the second part, I focus more specifically on the principles which will govern the model. To recap, I am aiming to construct a model for a candidate creative process, basing it upon an analogue of evolution by natural selection and setting it within a modified Fodorian massively modular structural framework. The case for assuming an extensive modular mental structure has already been examined. In this chapter, I investigate the proposed Darwinian analogy. Evolution occupies an at least superficially similar role to creativity, in that it generates novelty in nature. But this apparent correspondence does not, of course, necessarily imply that creativity is another instance of effectively the same process. Hence, the proposed analogy requires careful scrutiny before it can be adopted for use in the model. Kronfeldner's recent work *Darwinian Creativity and Memetics* (2011) incorporates an exposition of relevant aspects of Darwinism and a most thorough scrutiny of the case for creativity theories based on a Darwinian analogy. She reaches an essentially negative conclusion as to the analogy's explanatory power and potential, and has thus provided a strong and comprehensive counterargument against the intended line of approach. I will, however, undertake to demonstrate that a modified form of the analogy can be assumed as a basis for modelling without falling foul of the objections so raised.

The outcome of investigations in chapter two points towards the inclusion of a search and recognition element within a creative process, informed by an understanding of required properties. But these are features absent from biological evolution by natural selection, so they are by no means Darwinian, and might not even be compatible with Darwinism. It might therefore be queried as to whether they are instead indicative of a different type of process entirely (in which case I may need to check whether the basic

Darwinian pattern might not on its own be sufficient for creativity, or whether it should be abandoned in favour of some as yet unidentified alternative having these features).

Kronfeldner carefully examines Darwinian theories of creativity which are based on unguided, or undirected, variation, by analogy with evolution by natural selection, and concludes that these are inadequate in terms of explaining creativity. She finds that creativity calls for guided variation, which is not inconsistent with my own initial line of thinking, as above. But she does note some underlying similarity, in that creativity does at least conform to the basic variation and selection pattern, as found in Darwinian evolution. These findings do not conflict with my own view, that a modification of the Darwinian scheme might be appropriate, although they are not sufficient in themselves to confirm it.

Furthermore, I find that Kronfeldner's analysis is helpful in bringing the differences between Darwinian evolution and creativity into sharper focus. I intend to show that these are significant and that besides weakening the case for a strict Darwinian analogy, they also strengthen the case for a modified one, and tend to be indicative of features which will need to be included in any model.

Kronfeldner argues convincingly that selection in cultural innovation is not a single step function, and involves two separate stages (I would say, sometimes three – to bring in external pre-testing). This is a divergence from the single step selection characteristic of biological evolution, and which I will argue affords creativity more potential for enhancing efficiency and reducing risk relative to the basic Darwinian scheme.

I will support Kronfeldner's inclusion of (active) response to challenge among the characteristics of creativity, of which the added value attached to creative output is effectively a reflection. In my own conjecture it is the appreciation, and analysis, of such challenge which informs the requirement, the set of desirable properties sought in a creative product. The desirable properties will also form the basis of the recognition

of promising novel variants, so that the variation (resulting from the search) is influenced, or driven, by what will also be the selective factors.

The foundation for the analysis and the formulation of a requirement is the prior knowledge acquired by the creative individual or agent. This does not conflict with Kronfeldner's originality and spontaneity conditions, which refer to the relationship between prior knowledge and creative novelty. Properties desired can be known in advance, without compromising the novelty of creative output, whereas if the innovative element of the output itself were to be known beforehand, it would not be new. The role of prior information and knowledge in creativity differs from that of the accumulated effects of past change in biological evolution. As Kronfeldner points out, in creativity prior knowledge can be brought to bear on the assessment of the challenges presented by the current environment, whereas the cumulative effects of earlier changes in biological evolution refer back to previous environments. I relate spontaneity to searching and possibility space, in that novel combinations may be identified via searches. The use of prior knowledge to restrict variation mitigates the effects of the *Universal Library* problem.

In the course of this discussion, I begin to refine the outline process built up in earlier chapters. The modelling exercise itself will follow in chapter six.

5.2 Aims for modelling

From above, and following on from earlier chapters, the objectives for attempting to model a creative process can be taken as follows:

1. To explore the possibility and practicability of mapping out a route for a naturalistic process that will generate output which might potentially qualify as 'creative'.

2. To seek a naturalistic explanation of creativity by putting together a plausible process capable of creative output.

Naturalism is the view that all events are in principle explainable by recourse to natural laws and processes, and also that they are not influenced by any forces external to the natural world. Thus, a naturalised creative process would have to exclude any intervention arising through metaphysical or inspirational elements (intervention would be beyond explanation). In other words, such a creative process must avoid any discontinuity in the causal chain. A materialistic model of creativity would be compatible with naturalism, although naturalism does not necessarily require an ontology that is restricted to the material alone.

Thus the successful construction of an adequate model for a creative process would demonstrate that creativity is possible within a naturalistic ontology, so that at least some instances of creativity are in principle explainable. It would not on its own establish that human creativity follows the model, neither would it show that intervention is never involved in any creative process (only one black swan event being needed to disprove any such claim). But it would confirm that neither intervention nor *ex nihilo* creation is a necessary factor in generating creative output.

Whether or not a successful model for a creative process corresponds to human creativity, it ought to provide evidence that the construction of an artificial creative process, operating along the same lines, might in principle be achievable.

5.2.1 Methodology

Boxology is a technique for representing the logical structure of a process graphically, as a set of boxes together with links to show the relationships between them (along similar lines to a flow chart or organisation chart). The boxology will take as a framework a sub-Fodorian modular structure, retaining some of Fodor's criteria for mental modules, although not all of them, as already discussed. Any resultant model will necessarily be restricted by time and resource constraints to a high level of aggregation.

The feasibility and modelling exercise in effect constitutes an extended thought experiment, aiming to discover to what extent such a process might be conceivable, and how much progress might be made towards its specification.

The feasibility and modelling exercise in effect constitutes an extended thought experiment, aiming to discover to what extent such a process might be conceivable, and how much progress might be made towards its specification.

5.2.2 Benchmarks

There will be a need to assess the validity and robustness of the outcome of the modelling exercise. Appropriate means for establishing soundness will include the application of the following criteria:

1. whether the information flows and chain of causality are fully traceable from input to output, at the relatively high level of the model.

2. whether the proposed structure sits easily with commonly accepted definitions of modules and their interactions;

3. that no metaphysical or inspirational element is required at any point.

4. whether the model corresponds to a system that can be expected to generate output that meets adequate criteria for creative products.

On the final point, novelty and value are commonly held to be basic attributes (as discussed in chapter two), and thus the most significant indicators, of creativity. Therefore it seems reasonable to suggest that any candidate model must output products capable of satisfying these criteria at least. Ideally, it should qualify for some of the further criteria that have been suggested (as discussed in chapter two), such as Boden's stipulation that creativity should be 'surprising'.

5.2.3 Practical, realistically attainable expectations for outcome

Possible levels of success might be expected to include:

1. A plausible creative process is identified or constructed which clearly reflects the human creative process (or some instances of it): it imitates mental activity and thus contributes towards an explanation.

2. A plausible creative process is identified, capable of producing appropriate output, but not necessarily reflecting human creativity.

3. A plausible creative process is identified, capable of producing appropriate output, but clearly different from human creativity. (Such an outcome might be of interest in the context of artificial creativity.)

4. No specific process is identified, but there is some indication that such a process is achievable.

5. It is concluded that no explanation seems achievable.

6. No result, either positive or negative, is found.

The investigation will address the question as to how much progress can be made towards any of the first four outcomes. They would have different implications for overall findings:

1. To identify a process replicating human creativity would solve the mystery of creativity. In so doing, it would also resolve the *Meno* paradox. This degree of success is, of course, unlikely to be attainable in practice in the current study, given the time and resource constraints.

2. To identify a process satisfying some of the criteria for creative output but not necessarily corresponding to human creativity would constitute a naturalisation, but without demonstrating that all creativity is capable of naturalistic

explanation. Human creativity might be exclusively inspirational or metaphysical, even if other forms of creativity were possible and were not necessarily so.

3. To find no explanation to be achievable might provide a measure of support for the acceptance of inspirational theories of creativity, although that would depend on the reasons for such a conclusion (for example, the search may have to be abandoned if/in the event that it had become clear that any creative process would be far too complex for modelling to be practicable).

4. If no results were forthcoming, this would of course mean that no direct progress had been made in the direction of modelling and explanation, but it might at least close off some lines of investigation found to be unfruitful.

5.2.4 The basis for an assessment of feasibility

The foundations for an assessment of the feasibility of modelling a creative process are now in place, in outline at least.

Boxology provides a graphical method for high-level process modelling. For this to be successful, the modelled process will need to be intelligible, but not unfathomably complex (as Fodorian central systems might be) at least at the highest level of aggregation. The process also has to be reducible to stages in an unbroken causal chain, in other words, it must be capable of explanation.

The process will require some form of structured architecture against which it can operate. As above, a modified Fodorian modular (or massively modular) framework will be assumed. Modularity is by no means universally accepted, but the initial exploration in earlier chapters suggests that it might have appropriate characteristics to support the process modelling. If a plausible candidate process is identified and found capable of creative output, it may then be possible to discard, or transcend, some of the initial assumptions.

It will be assumed that evolution by natural selection can be taken as a sufficiently good analogy for creative process, *for the specific purpose of constructing a model*. The soundness or otherwise of this assumption for its intended role will be discussed later in this chapter. From earlier chapters we have the starting point that creativity involves bringing into existence something that was not there before, almost as if something had been obtained for nothing. The model has to demonstrate a feasible way of generating cultural novelty and value from a naturalistic, reproducible process. Evolution is a well-established and understood mechanism that performs an at least superficially similar function in nature, as further discussed below.

5.3 Stepping through a creative process

The assessment can begin by taking up the indications from earlier chapters as to the way in which a creative process might be broken down into distinct phases. Wallas's four stages of preparation, incubation, illumination, and verification have earned wide acceptance (as mentioned in chapter one). The discussion at the end of chapter two points towards a more extensive list, which can be set out as follows:

1. Acquisition of prior knowledge.
2. Assembly of a requirement (for a creative product with a particular set of properties).
3. A search for candidate ideas to fulfil the requirement.
4. Matching output from (3) to the requirement from (2).
5. Recognition (the successful matching of search outcome with requirement).
6. Evaluation of recognised matches by testing the outcomes.

This extended version maps roughly onto Wallas's stages: the acquisition of prior knowledge corresponds to Wallas's preparation stage, the search and matching process (which is not necessarily conscious) to incubation, recognition to illumination,

and evaluation to verification. The main difference from Wallas lies in the expansion of the incubation stage to include search and matching.

A seventh phase can be added over and above the core process covered by Wallas, to take account of the dissemination and preservation of cultural change through testimony, oral tradition, written records, data archives, and so on.

More interestingly, the extended process can also be mapped approximately onto a simplified sketch of evolutionary change, as in the following table.

Presumed phase of creative process (extended)	**Steps in evolutionary change**
1. Prior knowledge acquisition	Accumulated outcomes of previous evolutionary change
2. Understanding the requirement	Vacant or newly opened ecological niche in the environment
3. Search	Natural variation and mutation
4. Matching	Variants arise and have differing degrees of adaptedness relative to the niche
5. Recognition	A new variant arises which is capable of successfully taking advantage of the niche
6. Evaluation	The new variant is tested by environmental challenge
7. Cultural preservation	Genetic transmission of the new variation down through subsequent generations

5.3.1 Evolution by natural selection as an analogy for human creativity

The case for taking evolution as an analogy for human creativity has been mentioned in earlier chapters, and can now be summarised here.

Evolutionary change produces novelty in nature. It can give rise to increased complexity and emergent properties, so that more can be generated than was present in the input raw materials at the start. Arising novelty is tested and only those variants adapted to (in effect, of value in) the environment are preserved. This is the closest approach in nature to the function of creativity in generating novelty and value in culture, and therefore evolution is a strong candidate analogy for creativity. However,

evolution does not require the intervention of a designer; 'design' arises through evolutionary processes operating according to laws of nature, in a manner conducive to explanation.

All this cannot be taken to imply that creativity has any direct equivalence to evolutionary change. In spite of the mapping above, the correspondence would not be expected to be exact. On the face of it, human creativity would appear to have certain inbuilt advantages over evolution, arising from its being embedded in a psychological, as opposed to physical, ontological setting. In natural selection, the benefits of the cumulative results of previous change have accrued in current conditions, and also those of past environments. But successful accommodation to external change requires the generation of variants adapted to new environments and to challenges to be presented in the future. The accumulated evolutionary design elements can be expected to confer advantage, but genetic material has no means of prediction. Humans have the facility to make forecasts of future needs and conditions, to extrapolate from current circumstances and imagine future possible events. These are subject to error and the unforeseen, of course, but permit the direction of effort in pursuit of anticipated future reward. New ideas do not tend to turn up to order, and might be serendipitous, but relevant background information can be put together with intention and purpose[38].

These differences and advantages attach to creativity due to its not being a straightforwardly physical process in nature, operating under natural laws. They lend at least superficial support to attempts to look for the origins of human creativity on the basis of evolutionary psychology. The initial similarities and differences between creativity and evolution by natural selection become clearer on examining the extended list of phases in more detail.

1. Acquisition of prior knowledge

[38] Perkins (1981, pp.100-101) notes the importance of purpose.

Prior knowledge is the equivalent of the accumulated results of previous evolution, according to the proposed analogy. Its acquisition is an essential foundation for creative thought, just as Kekulé formulated a structure for the benzene molecule following much previous work on organic compounds.

As said above, prior learning can be purposeful, and centred on perceived future needs or areas of interest as well as current and past ones, whereas DNA only encodes the outcomes of earlier evolutionary changes, prompted by the challenges presented by ancestral environments.

Accumulated evolutionary change is also a constraint, in that development in one direction in response to current conditions can close off other options, which might have become beneficial under different circumstances at some point in the future. Prior knowledge can likewise be a constraining factor on creativity, because it forms part of a world picture, a set of beliefs that may need to be abandoned or adjusted to accommodate new ideas. Freedom from preconceptions is favourable to creative thinking, so that there is a trade-off between the benefits of accepting current theories as groundwork for creative thought and their drawbacks in terms of inertia (because discarding or adjusting previous belief sets can be costly and inconvenient).

2. Assembly of a requirement (for a creative product with a particular set of properties)

Salient prior knowledge is essential for framing a *requirement* (that is, a specification of the properties sought), which can only be put together on the basis of what is already known. In order to become aware that benefits could be obtained from the implementation of a particular set of properties (if only means could be identified by which to bring this about) sufficient knowledge will be needed of the area concerned, as in the cases of Kekulé and Fleming, who were in possession of the necessary experience in their respective fields.

It is this second phase, setting up the focus provided by the requirement, that sidesteps the *Meno* paradox. Given the appropriate prior knowledge it may be possible to specify the set of properties required to fulfil a particular need, even if there is nothing known that will actually possess those properties. The properties might then be recognised, even if the entity embodying them might not be. Fleming was aware that there was a need for something that would kill bacteria, and could therefore recognise the bactericidal properties of penicillin when by chance they became evident, even though he was not looking for penicillin itself, and indeed did know that it would be helpful to seek it, exactly as the *Meno* paradox anticipates.

Acknowledging the role of the requirement thus avoids the *Meno* problem; it is not known where to look or what to look for, but the required characteristics *are* known. This is a point of difference from evolution, where the analogue of the requirement is the existence of the ecological niche. A requirement can drive an active search, whereas the natural variation and mutation that might lead to adaptation to the environmental niche is random, or at least passive.

The role of environmental influences was briefly discussed in chapter one. In effect, environmental challenge - such as climate change - defines an (analogue) requirement for adaptive variants in nature.

3. A search for candidate ideas

Searching can be directed: a search can be *for* something *in* a particular place. In the case of human creativity, it is suggested that a search might be driven by a requirement for a specified set of properties, formulated as above. Direction holds the potential for improved efficiency; this is another point of difference from the proposed evolutionary analogue. The preserved results of previous evolutionary change may assist in reducing the (otherwise vast) scope of natural variation but that restriction may also act as a constraint, as above.

There are two target zones that might be searched in a creative process, because the solutions, or partial solutions, might be found either in stored memory or in the external world, via sense data. Directed search implies structured internal storage. The information would have to be accessible, to be sufficiently organised that searching can be carried out efficiently. This would be the equivalent of indexing a subset of the *Universal Library*, as discussed in chapter two, and would permit the scope of a search to be confined to within feasible bounds.

External sense data are clearly not structured; search may be directed to promising locations (such as a library or science laboratory) although here there is a place for random or coincidental observations as in serendipitous discovery (perhaps this might be taken as equivalent to mutation in an evolutionary analogue).

Searching is of course found in nature. Hunting for prey, viewed as a search, is a situation not dissimilar to serendipitous discovery through observation of the external environment.

Internal searching of stored memory information might be seen as more characteristically human; perhaps this type of search only comes into prominence with the development of language and concepts, offering opportunities for structuring, as will be discussed later on[39].

4. Matching search outcomes (from 3) to the requirement (from 2)

In any creative process involving a requirement-driven internal search, there will need to be some means of matching, of comparing the findings from the search with the requirement to determine the goodness of fit. This will be explored further in the next chapter. The creative use of external source material depends upon understanding the significance of sense data and then matching the interpreted information with known requirements.

[39] See also the discussion in chapter two.

5. Recognition (of a candidate for a successful match)

Perkins refers to the importance of noticing and recognition (see 1981, pp.76-87). In the current scheme, the recognition of a potentially creative thought is equivalent to the successful match of a search result with a known requirement. It seems reasonable to assume that the match can generally be recognised (and mentally evaluated), more quickly and efficiently than the proposed evolutionary analogue, the occupation of a new or vacant ecological niche by a successful new variant.

6. Evaluation of recognised matches by testing the outcomes

The recognition of a successful match of search outcome with the required set of properties provides a candidate novel idea. But added value has to be established, as well as novelty. This requires some form of testing, to determine whether the occurrence of the required properties might be expected to perform adequately to merit adoption or further investigation.

Evaluation in a creative process has a considerable advantage over the proposed evolutionary analogue. New approaches can be tried out in possibility space, by setting up possible worlds in the imagination in which the change is implemented[40]. This allows initial testing to be performed without incurring the risk which accompanies trial in the actual world or immediate exposure to the wider culture. Mental simulation is painless, whereas a maladaptive genetic variant can be fatal for those individuals that have it.

7. Preservation of cultural change through tradition, testimony, records etc.

In human creativity, there is a second stage of testing (see the comments on Kronfeldner, below). A creative idea, once identified and confirmed by the agent, might afterwards be taken up by others. If it gains acceptance it might then be recorded and transmitted to other cultures or to subsequent generations.

[40] See the discussion in chapter three.

5.4 The proposed Darwinian analogy

As said, one advantage of taking Darwinian evolution as an analogue is that it can be explained in terms of the operation of the laws of nature. If creativity is in any sense a reflection of the same process, at another level of ontology, then this is very close to a naturalistic explanation of creativity. But even though the points of comparison between evolution and creativity highlighted in the above listing are suggestive, and bring out some correspondences, they do not by themselves add up to a demonstration of an adequate analogy. For current purposes, however, it is only necessary that the analogy should be sufficiently strong to support the construction of a model. In that respect, some of the following considerations may be of significance.

1. If evolution were to offer a complete and fully adequate analogy, then modelling a creative process should be at least theoretically achievable. The model would be intelligible, and should be equivalent to a naturalistic explanation of creativity. (However, as above, there are clear points of difference even on a superficial view.)

2. If evolution is a partial or incomplete analogue, and has substantial points of divergence, it might still be possible to set up a model for a creative process, taking evolution by natural selection as the initial basis for construction.

3. A model can only be constructed for a naturalistic process. A successful model, whether or not based on evolution, demonstrates that creative novelty can arise from a naturalistic process, even if the model is not applicable to specifically human creativity.

4. A successful model provides a candidate explanation, whether or not it is based on an analogy with evolution, and if it is, then it does not necessarily destroy the explanation if the analogy turns out to be false, or the logical basis for the analogy is subsequently found to be mistaken.

This project is particularly concerned with the origin of creative novelty; Darwinian evolution is at least superficially similar to creativity in that evolution by natural selection is a mechanism, or process, that outputs novelty - successful, adaptive novelty. Evolution by natural selection explains the origin of organic novelty by showing how it is brought about. It is therefore the obvious place to look for a template for creativity, itself a generator of novelty and added value. I have already discussed similarities and differences between the two, and although there is some underlying commonality, there are also significant points of divergence. Both of them originate novelty. The basic novelty and value criteria for creative products recall the variation and selection stages in the process of evolution by natural selection. But I have noted a number of indications that a strict, straightforward analogy will not be adequate, the chief of which is that human creativity is a response to challenge as opposed to a result of (environmental) challenge, as is the case for biological evolution.

In human creativity, the properties desirable in a creative product can be known in advance, on the basis of prior knowledge of the field concerned. The required properties might in turn be expected to enable an active search of internal and external data, focused on the properties in question, rather than directly on the solutions that might embody them. Humans can search for required properties and can recognise the properties when found (by comparing the retrieved set of properties with those required). On the other hand, evolution by natural selection can have no requirement, search, or recognition, because there is nothing, no agency, that can in any sense (analogue) understand, analyse, or appreciate the nature of environmental demands. I have also noted observations that seem to hint at the structured data storage that would facilitate searching; this would permit relevant prior knowledge to be organised to promote targeted access. Creativity appears to be in some way dependent on prior knowledge, which is not available to natural selection. Instead, in organic evolution information about past evolutionary change is recorded at molecular level (but this is

not equivalent to prior information on the current environment[41]). These findings have helped to clarify the vision of the type of process which might mediate creative thinking, that is, a process which shares some of the characteristics of evolution by natural selection, without being a straightforward analogy.

Darwinism involves natural selection. Artificial selection arises from human intervention and might perhaps be viewed as a kind of halfway house between evolution and creativity. It represents an attempt to produce a variant with desirable properties (organisms tailored to human requirements), which can be established in advance. The properties desirable in a creative product might be appreciated beforehand in a similar fashion.

This section is about the *process* – the question as to *how* exactly does evolution by natural selection work, and to what extent could creativity be replicated by a process operating in the same way? To this end, Kronfeldner has already published a commendably thorough critical analysis of Darwinian theories of creativity, in terms of the origination of novelty, with essentially negative conclusions. I intend to show that it is quite possible to accept all her criticisms without destroying the case for taking a – suitably modified – analogy as a foundation for modelling a creative process, because the criticisms are often pointers to essential differences between the two processes, which exist alongside the similarities.

The extended list of phases set out above, together with the approximate mapping onto evolution by natural selection, gives a potential starting point for modelling, but it is necessary to check for possible flaws and objections, and counterarguments. Kronfeldner's much more rigorous analysis of the case for taking evolution as an analogy for creativity, in the earlier chapters of *Darwinian Creativity and Memetics* (2011), is particularly helpful in providing a counterargument to a straightforward Darwinian analogy, and in pointing out why this is inadequate as an explanation of

[41] As Kronfeldner points out, see later in this chapter.

creativity. Therefore I will examine parts of this account in some detail (though it will need to be borne in mind throughout that Kronfeldner is not looking at the analogy with a view to using it as a basis for a model).

I will commence with a summary of a few salient points selected from Kronfeldner's investigation, and then follow this up with comment and discussion.

5.4.1 Kronfeldner's analysis

As discussed earlier, the most widely accepted characteristics of creativity are novelty and value, and Kronfeldner's description[42] includes both of these, relating them to the environment and response to challenge.

Originality and spontaneity

Kronfeldner also refers to criteria concerning originality and spontaneity (2011, p37). The originality condition demands at least partial independence from any original, so that creativity and learning are in 'gradual opposition' (2011, p.38). They are not totally opposed, because for example Kekulé could not have formulated his ring structure for the benzene molecule without prior knowledge of chemistry. But this prior knowledge, although necessary, was insufficient on its own (2011, pp.37-38). Kronfeldner further argues that creativity requires spontaneity in addition to originality, which positions it in another gradual opposition, this time to routine (mechanical) production. There must be at least an element that is independent of the creative individual's prior knowledge. (2011, p.38). She adds that Plato's *Meno* paradox implies spontaneity because 'you cannot search for or intend to bring about what you do not know ... even if you know to some extent the criteria that have to be fulfilled…' (2011, p.38). She writes:

> On this basis, the characteristic passivity involved in creativity, of which Picasso spoke, can be interpreted as a peculiar simultaneity: you only know that you looked for *it* at the very moment of insight. (2011, p.38).

The four stages

[42] See chapter two.

Kronfeldner notes that the gradual opposition to routine methods manifests in serendipity, trial-and-error, and in flashes of sudden illumination or insight (as for Kekulé). She refers to Poincaré's[43] description of four stages as widely accepted in the psychology of creativity, naming them as: preparation; incubation[44]; inspiration; and elaboration and evaluation (2011, p.39-40, and see also the discussion on Wallas in chapter one, above). She observes that the incubation and inspiration stages bear most directly on the spontaneity of creative thought or insight, and that they imply passivity on the part of the creative person. (2011, p.40). She concludes that the call for ingredients of spontaneity and originality carries the implication 'that creativity cannot fully be *explained* by reference to learning and previously acquired knowledge alone.' (2011, p.40).

Darwinian theories

Darwinian approaches regard cultural change as an evolutionary process, based on an analogy taken from nature. Kronfeldner explains that in Darwinian evolution variation and selection are essential ingredients of the mechanism of change. Given variation between individuals, evolutionary change of the whole system is driven by some form of 'sorting process' (such as natural or artificial selection) (2011, pp.16-17). This selective component alters the relative frequencies of occurrence of the variants in the population (2011, p.16).

Kronfeldner links creativity with a particular variety of Darwinian analogy, relating to 'the *origination* of novelty'[45] (2011, p.viii). In sum, she does not find in favour of the analogy, concluding:

> … I illustrated that creativity is not based on undirected variation, even though culture itself can be regarded as a gradually changing variational system based on selection by groups of individuals. First, novelty in culture is already pre-selected by the creative individual when it enters the cultural pool. Second,

[43] See Poincaré (1908).
[44] Incubation can coincide with a temporary disengagement from the question concerned.
[45] Kronfeldner identifies three underlying types of Darwinian analogy (2011, p.viii), of which this one is particularly relevant for current purposes, and centres on the question of origination.

> there is blind variation in the sense that there is no guarantee that the ideas produced by creative individuals are appropriate. Creativity is in this unfortunately trivial sense Darwinian. It is, however, not Darwinian in a stronger sense since it is based on guided variation. (2011, p.135).

Kronfeldner effectively specifies two ontological categories, (biological) nature and culture: 'Life constantly changes and so does culture.' (2011, p.vii) She thus implicitly treats the world as being split into two ontologically distinct realms corresponding to nature and culture, from the viewpoint of change, as has already been discussed in chapter two. She points out that there are currently different theories of the origination of novelty corresponding to each category; she draws attention to the contrast between the general, folk assumption that cultural change is the outcome of design, and the absence of any call for input from a designer to evolution in nature.

In the above extract, Kronfeldner also identifies a two-phase scheme of selection at the cultural level: internal pre-selection by the creative agent, with subsequent further selection by others within the cultural setting. Earlier on, she notes that at the cultural level there is constant change: there is variation, in that individuals within the culture have different ideas, and innovation is sporadic and locally initiated; also, there is selection due to individual choices[46] (2011, pp.42-44).

Kronfeldner traces the Darwinian approaches to culture back to an origin in two papers published in the 1960s by Campbell[47] and notes that he, and also Simonton[48], view 'creativity as based on a Darwinian process of "blind" trial and selection.' (2011, p.3). She emphasises the need for clarity as to the nature of the 'blindness' here implied (2011, pp.3-4). She notes that:

> Today it is quite common to say that Darwinism rests on blind, random or undirected variation... The problem is that it is often unclear what blindness (i.e. undirectedness), means exactly (2011, p.17).

Blind variation

[46] (To which I would add institutional and political influences.)
[47] Kronfeldner cites Campbell (1960) – see also chapter two – together with Campbell (1974).
[48] Citing Simonton (1999), among other publications.

Kronfeldner quotes[49] Campbell's statement that: "[I]n going beyond what is already known, one cannot but go blindly. If one can go wisely, this indicates already achieved wisdom of some general sort" (2011, p.45). She relates this to the *Meno* paradox[50]; creative novelty cannot be derived directly from prior knowledge and still be counted as new. Kronfeldner considers three possible interpretations of 'blind', as referring to: random, unjustified, or undirected variation, and argues that the variation involved in creative thought process cannot be so completely random that all outcomes have the same probability, which is not true of the variation element in biological evolution, either (2011, pp.46-47). She says that 'blind' could be taken to mean unjustified, in that trials for novel ideas could not be justified in advance by the success of their result, which cannot be known beforehand. The possibility of unsuccessful trials cannot be eliminated in this situation, so that: '… creativity is *necessarily* selective and in this sense Darwinian.' (2011, p.49). She argues that if this is all that the analogy implies, then it does not offer any new understanding and is '*heuristically trivial*'. Also, unjustified variation merely reflects the originality and spontaneity conditions, and is thus '*explanatorily trivial*' as well (2011, pp.50-51).

Neo-Darwinism holds that in (biological) evolution by natural selection the selective factors are 'decoupled' (2011, p51) from those influencing variation, so that the variation is entirely undirected[51]; Kronfeldner argues that variation in creative process would have to be similarly decoupled, and variation blind in this stronger, undirected sense, to provide for a strict analogy between creativity and biological evolution. She states that the notion that creativity involves undirected variation has been criticised on the basis that humans innovate in a planned and purposeful way, with the benefit of accumulated knowledge and experience (and thus in a directed, or guided, fashion, unlike evolution by natural selection). She identifies three '*compatibility arguments*'

[49] Citing Campbell ([1974] 1987) in: Radnitzky and Bartley (eds.) (1987, p.57).
[50] (See chapter two.)
[51] In this situation, the generation of variants is isolated from factors involved in selection, so that 'variation and selection are "decoupled"', as opposed to "coupled", where variation and selection are correlated and hence variation is directed. (2011, pp.19-20, and citing Toulmin (1972)).

that acknowledge this criticism but respond that the analogy can still be maintained. Kronfeldner concludes that these arguments fail or amount to relatively 'trivial claims' (2011, p.53).

She discusses examples including those of Kekulé and Fleming. In Kekulé's case, she observes that prior knowledge of the background 'structures the search space and triggers certain ideas' (2011, p.56). That same set of prior knowledge also informs recognition of the solution, so that the selecting factors are not unconnected with those bearing on the variation (2011, p.57). On the other hand, Fleming discovered penicillin even though he was not actively looking for it at the time – this truly serendipitous discovery therefore has a random component, even though the recognition of the solution was enabled by prior knowledge (2011, pp.55-57).

One of the compatibility arguments, which she calls '*hidden chaos*' (2011, p.63) refers to a proposed 'pre-conscious or unconscious mechanism' (2011, p.63) for generating variation. The outcome then feeds into a conscious process involving guided variation. The unconscious generation of variants would be blind, as for biological evolution. She traces this notion back ultimately to Poincaré. She cites Simonton's (1988, 1995) "chance-configuration theory" as deriving from Campbell and Poincaré, and involving just such a hidden chaos (2011, p.65), which would produce large numbers of unconscious trials. The hidden chaos would correspond with the incubation and insight (illumination) stages of the Poincaré-Wallas four-stage model. Kronfeldner also identifies other theories as to the processes occurring at these stages, one being the '*opportunistic-assimilation hypothesis*' (2011, p.66) which regards insight as being explainable as the outcome of assimilation of - I would assume external - coincidental, serendipitous input. She cites Colleen Seifert *et al*. (1995) as showing that extending the length of time available for incubation did not increase the likelihood of the occurrence of insight (which would be expected to result under hidden chaos theories, where this ought to allow for greater numbers of idea-combinations to be produced).

Instead, the opportunity for influence by coincidences increased the probability of insight. Thus Kronfeldner concludes that this evidence does not support the hidden chaos notion.

Kronfeldner also refers to an '*ordinary-process view*', (2011, p.66) according to which such processes[52] are sufficient as explanation for creativity. These theories do not depend on any special mechanism for creativity. The ordinary processes involved are thought to operate rapidly, and not necessarily to be tracked at conscious level, which would explain why creative people cannot report on their operation.

Kronfeldner concludes that there is insufficient ground for claiming that creativity is based on undirected variation, in the absence of better evidence for any hidden chaos mechanism. She argues that creativity exhibits variation and selection, but that this contributes little in terms of new explanation. This much is also compatible with the ordinary-process view but the Darwinian theory of creativity therefore is not capable of replacing the ordinary-process view, which has begun to be taken up within the psychological study of creativity (2011, pp.67-74).

5.4.2 Discussion

This section contains my commentary on selected parts of Kronfeldner's investigation. I have included a few diagrams of my own as illustration.

If taken at face value, Kronfeldner's essentially negative conclusions might suggest that there would be no utility in pursuing Darwinian analogies for creativity any further, but the arguments concerned need to be viewed against the context of the aims of the current project. A Darwinian analogy, while not necessarily *per se* adding to an understanding of creativity, might still be sufficiently powerful to provide a structure

[52] (giving the following examples: '…perception, visual imagery, diverse cognitive heuristics (such as distorting, repeating, omitting and mixing parts of concepts and images), associational linkage, conceptual combination, analogical reasoning, abstraction, use of metaphors, conceptual expansion, memory retrieval, spreading activation, opportunistic assimilation, and so on. (2011, p.66)).

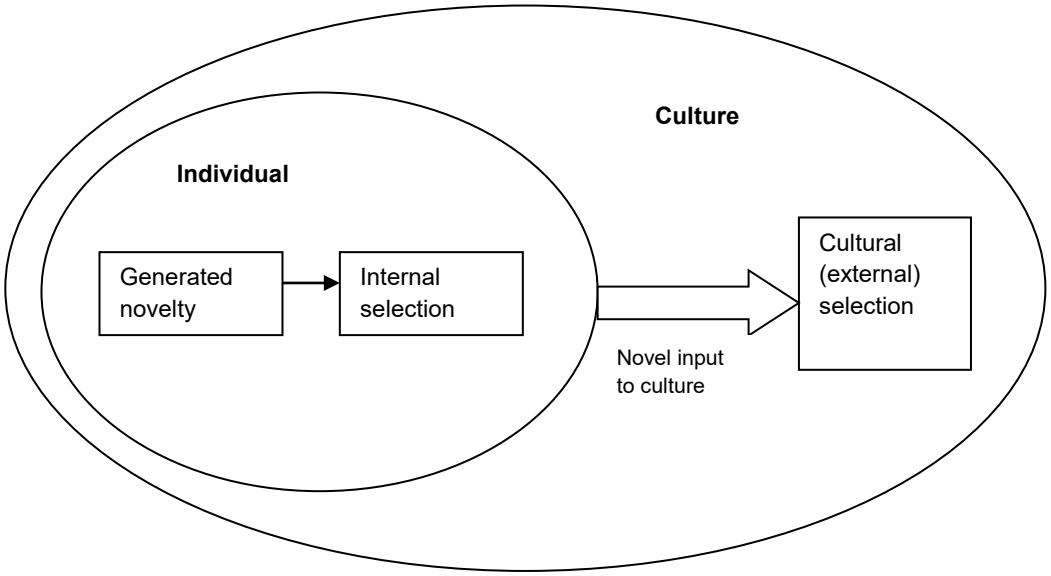
within which, or from which, a model for a creative process can be constructed. This remains so even if the analogy is only taken in an incomplete fashion, and there are indications, as outlined earlier, that a number of points of difference can be expected between Darwinian evolution and creative process.

In this context, Kronfeldner's precision in setting out the characteristics of Darwinian evolution and natural selection, and in examining any potential applicability of these to creativity and cultural change, is particularly to be welcomed and brings to light (or helps to clarify) a number of points of potential relevance which will need to be taken into consideration for modelling.

Two-phase selection

Kronfeldner's reference to internal pre-selection sets up a two-phase model of (cultural) selection; I have sketched such a configuration in my diagram below.

Fig. 5.1: Two-phase selection of cultural novelty



The two stages of selection as illustrated above are:

(1) within the individual (by the individual's own internal testing); and

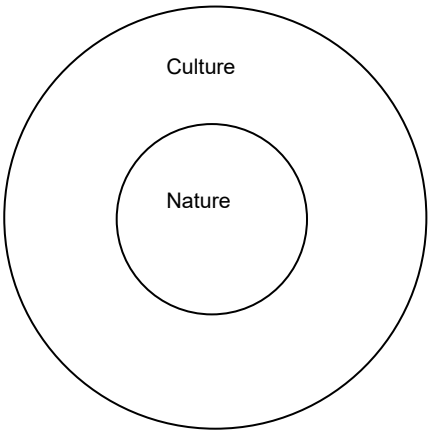
(2) outside the individual but within the culture.

Novelty is generated internally to the individual and if it passes their own internal selection testing it might then be chosen for implementation in, or exposure to, the external, cultural environment (it is, of course, the individual, cognitive phase which is of more interest for the current project, due to the emphasis on the origination of novelty).

I would also add a third, intermediate, phase to the diagram, corresponding to external pre-selection by experimental testing, for example by carrying out virtual trials in cyberspace, on a sketch pad, in a laboratory, and so on. This all points to a divergence from the biological pattern, as neither the two-phase nor three-phase versions would apply to Darwinian evolution, where the sole selective agent is the challenge provided by the external environment.

Darwinian views of cultural development

Fig. 5.2: Culture and nature



In this diagram, I have constructed a simple sketch to represent a division of the cosmos along the nature/culture faultline, so that it may be compared with the similar diagrams presented to accompany the discussion of Mill's inclusive and exclusive versions of nature, in chapter two.

According to Kronfeldner, creativity is a response to internal or external (environmental) challenge[53]. I would regard appropriateness to environmental conditions as being closely related (if not directly equivalent) to the potential of a creative product to contribute added value. I have taken this to be indicative of a key parallel with organic evolution, as both are at least partially driven by environmental factors, but, as discussed earlier, the active phrasing also points to a significant divergence between the two systems.

Organic evolution, Kronfeldner notes (2011, p.2), is insufficient by itself to explain change in culture, which is treated as an independent system in this respect. Perhaps another way of looking at this would be to regard the biological layer as providing a

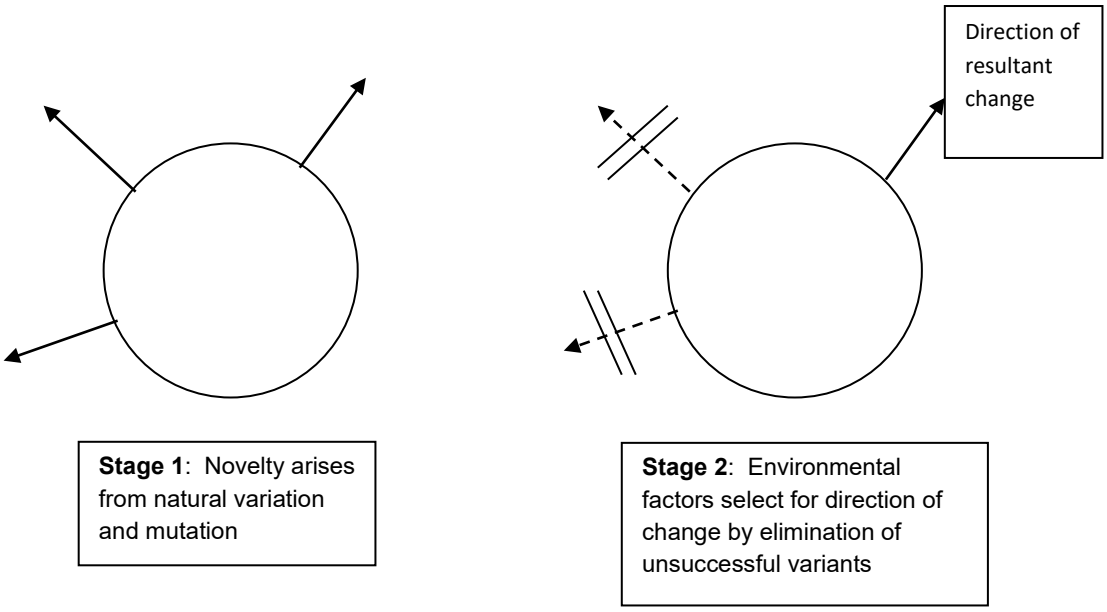
[53] See chapter two, above.

platform upon which culture is free to develop, just as, in computing, hardware only to an extent constrains the software that can be run on it.

On variation and selection

In biological evolution, these two components correspond to novelty and adaptedness: variation produces novelty, selection eliminates non-adaptive change. In this, there is already some *prima facie* similarity with novelty and value in creativity. Natural selection determines the direction of change, as Kronfeldner says; this can be sketched as in the diagram below:

Fig. 5.3: Natural selection (after Darwin); viewed as a two-stage process



There are two points of particular note here:

1. Selection follows the appearance of novelty.

2. It is selection that drives evolution, by dictating the direction it should take.

There is a clear-cut division between the effects of the two stages. The novelty is originated as a new variant, in the variation stage. The direction of change is determined at the selection stage, by the differential effects of environmental challenge factors on the types of variant generated in the first stage.

Variation and directedness

This prompts the question as to whether, or to what extent, the variation in a creative process - the origination of novelty – is itself directed, and if so, what is the source of the direction? I have already discussed, as a potential difference between the evolutionary and creative processes, two factors that might influence variation in the generation of creative novelty (if they do, in fact, apply). In understanding the required properties for a creative product, a human creative agent acquires the necessary prior information to direct a search for those properties. It is the search that is directed, not

the variation, but the makeup of the search criteria will clearly influence the output. Similarly, if prior information holdings (as the target of the search) are structured, then the characteristics of the structure will also have an effect on the outcome of the search.

If a creative process were to be taken as directly analogous to biological evolution by natural selection, then compliance with the principle of undirectedness would require a pathway generating novelty, but without being influenced by selective factors – this would clearly not be the case for any model dependent upon a requirement-driven search, as described above.

In biological evolution, some degree of restriction on the scope of variation might be expected to be essential to allow for any evolutionary change to occur at all, given the potential numbers of possible combinations and the *Universal Library* problem, as discussed in chapter two, and Dennett's *Darwin's Dangerous Idea* (1995, pp107-113). This is a particularly relevant issue for the current project. Kronfeldner notes that:

developmental constraints are (if at all) the effects of *past* selective environments. Directed variation (i.e. coupling) […] would be a direct effect of the *current* selective environment. (2011, p.21).

If direction is introduced into the variation aspect of a creative process due to a requirement driven search, then that will derive from prior knowledge about the current environment (as would be the case in directed evolutionary variation) rather than from the effect of past environments, as for evolutionary developmental constraints.

Fitness, value, and the potential for tautology

Here, 'value' might be suggested as the analogue of 'fitness', in a creative process – see also the discussion of the context-sensitive nature of the value criterion, in chapter two above. A corresponding tautology might then be suspected if the value of a novel idea were to be assessed according to the extent of its eventual distribution within the

culture, or perhaps by its price (for art, for example) or benefits in monetary terms (for a technological innovation). A creative idea might fail to spread, or to be adopted within the culture, even though it could have brought added value (it might perhaps be suppressed for political reasons, among others). Thus, in the much same way as for 'fitness', value is ideally thought of as potential, rather than as realised gain. Perceived added value explains why an idea might come to be adopted in the culture, although lack of value is not always the reason why a similar idea might have been ignored. But, as said previously, value is in any case contextually based, so for instance an artistic style may be unpopular at one time but then become appreciated and sought after a few decades later; there is therefore no way to ascribe an absolute measure of the potential value to a novel idea, and to an extent the tautology reflects this difficulty.

Originality, spontaneity, and passivity

In sum, in Kronfeldner's view, the originality and spontaneity conditions correspond to two forms of gradual opposition. Creativity is gradually opposed to:

(i) Learning or copying (originality condition);

(ii) Routine or mechanical production (spontaneity condition).

I would regard both of these conditions as essentially being qualifications of the basic novelty criterion for creativity; they refer to the character of the novelty that is generated.

Even Kekulé's specialist experience with carbon compounds would not alone have provided the solution he sought for the molecular configuration of benzene, because the closed loop structure was previously unknown. Kekulé had to supply this crucial additional ingredient from elsewhere. This new (and original) component could not be derived from current understanding of the field, but I would suggest that it is the kind of material that might be retrieved by a search for the required properties (alternative

structures that might be able to accommodate the composition of benzene given its known valency characteristics).

When spontaneity and passivity (to which Kronfeldner also alludes) are viewed in terms of modularity, they then become suggestive of the kind of effect that it might be anticipated would result from encapsulation. I would expect the individual concerned to be unaware of the operation of an encapsulated process while it was in progress, because the encapsulation would isolate the procedure from the rest of the mind. The output might therefore appear to be disconnected from any inputs of which the individual might have been aware, and thus to have arisen in just such a spontaneous manner. (Also, encapsulation might provoke experience superficially similar to that of the 'sleep of reason' – see chapter one).

The Meno *paradox and required properties*

The 'criteria … to be fulfilled', to which Kronfeldner refers[54], approximate to the *requirement* which a creative product is sought to meet. I discussed the *Meno* paradox earlier on, in chapter two, and there I suggested that a known requirement could in turn drive a search (but for the required properties, not directly for an entity or solution in which they might be manifested, as the *Meno* has it). The required properties do not in themselves entail any particular solution, and so the solution cannot be derived from the requirement alone. The *Meno* problem presumably also helps to explain why it might seem as if a creatively novel idea comes from nowhere, or nothing, as it cannot be known before the actual moment of realisation itself, otherwise it would not be new[55]. Thus, in a sense, if an idea is not spontaneous, it cannot be new at all.

Here, I might conjecture that perhaps – short of supernatural inspiration – a creative individual cannot create something that is absent from, or beyond, their personal possibility space, the total of all possible combinations of their current holdings of

[54] Kronfeldner (2011, p.38), and see earlier in this chapter.
[55] (This might be compared with the ontological event that is under way at the moment of movement from possible to actual.)

elements of information (memory, knowledge, belief, images) unless serendipitous observation is involved. If so, serendipity has something in common with inspiration, as constituting an external influence. Also, if novel internal combinations result from requirement-driven, directed search – as suggested above – then serendipitous observation retains a random element.

This links with the notion of the source for creative ideas, as discussed in earlier chapters. On the same hypothesis, Kekulé would have found both carbon chemistry (as then understood) and closed curves within his prior knowledge, and so the combination of the two (the ring structure of the benzene molecule) even though entirely novel, lay within his personal possibility space. Similarly, Fleming could combine his observation of the action of penicillin mould (in the form of an internal, mental representation) with his prior knowledge of bacteriology, so that his serendipitous observation brought the novel possibility of the application of this property of the mould within his personal possibility space.

In each case, the new idea would have been absent from the creative individual's prior knowledge of the topic area, and so could not be deduced on the basis of this alone. Instead, the novel combination is formed by stepping beyond the topic area itself and drawing an element from another domain of thought. The combination is thus somewhere within the creative individual's (vast) possibility space, even though the connection had not previously been made.

In this way, the realisation of the possibility of the novel combination might be expected to have the appearance of spontaneity, exactly as Kronfeldner describes. It was not in the individual's prior knowledge at all, but it was within scope of their personal possibility space, only coming to light when the appropriate combination was identified.

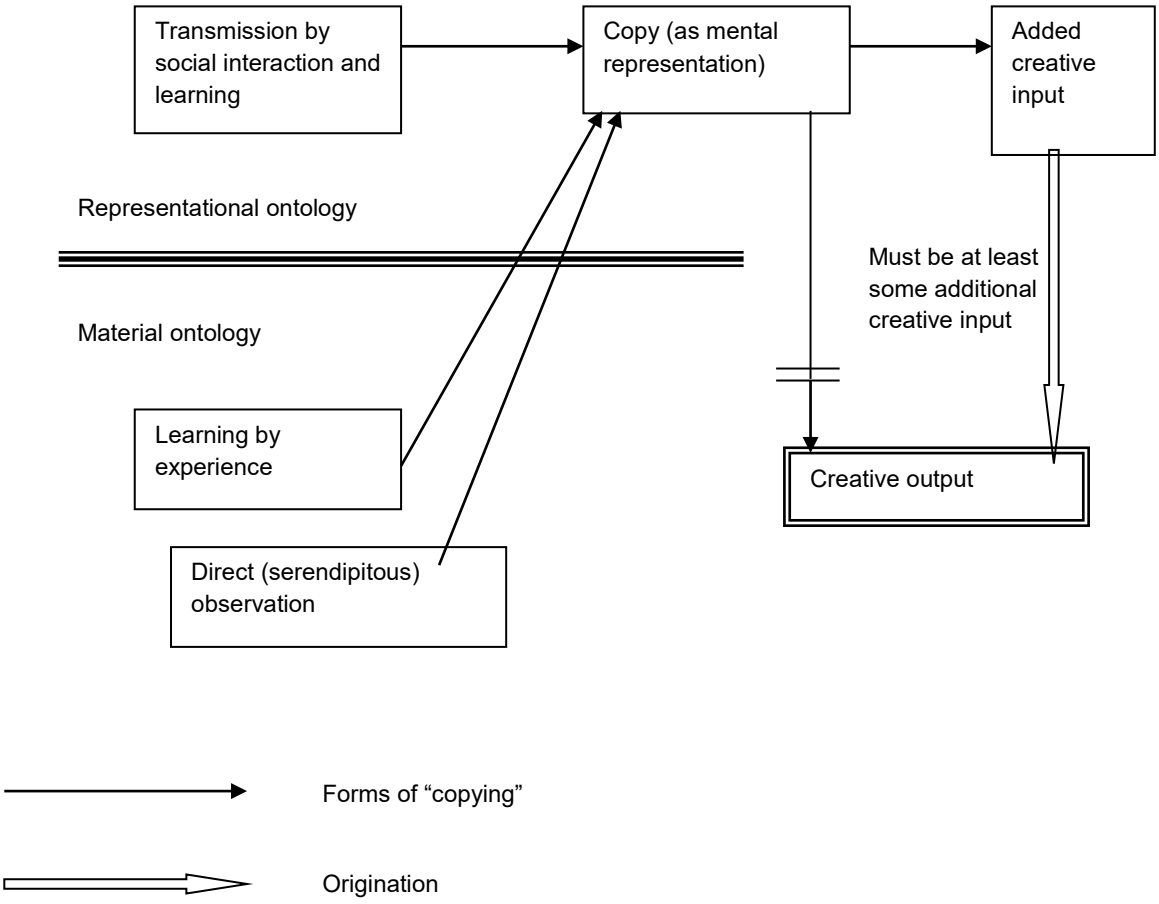
On spontaneity, it is interesting to compare the following passage from John G. Bennett's *Creative Thinking:*

> There is the saying that creativity is ninety-nine parts perspiration and one part inspiration – in other words, that it is mostly very hard work. But there is no doubt of the importance of some second element in creativity that we ourselves cannot control. I am going to call that element spontaneity. If you read accounts of creative activity by scientists, artists, or others, you see in these accounts that the spontaneous element is really out of the person's control. (1998, p1).

This again recalls the character of an encapsulated process.

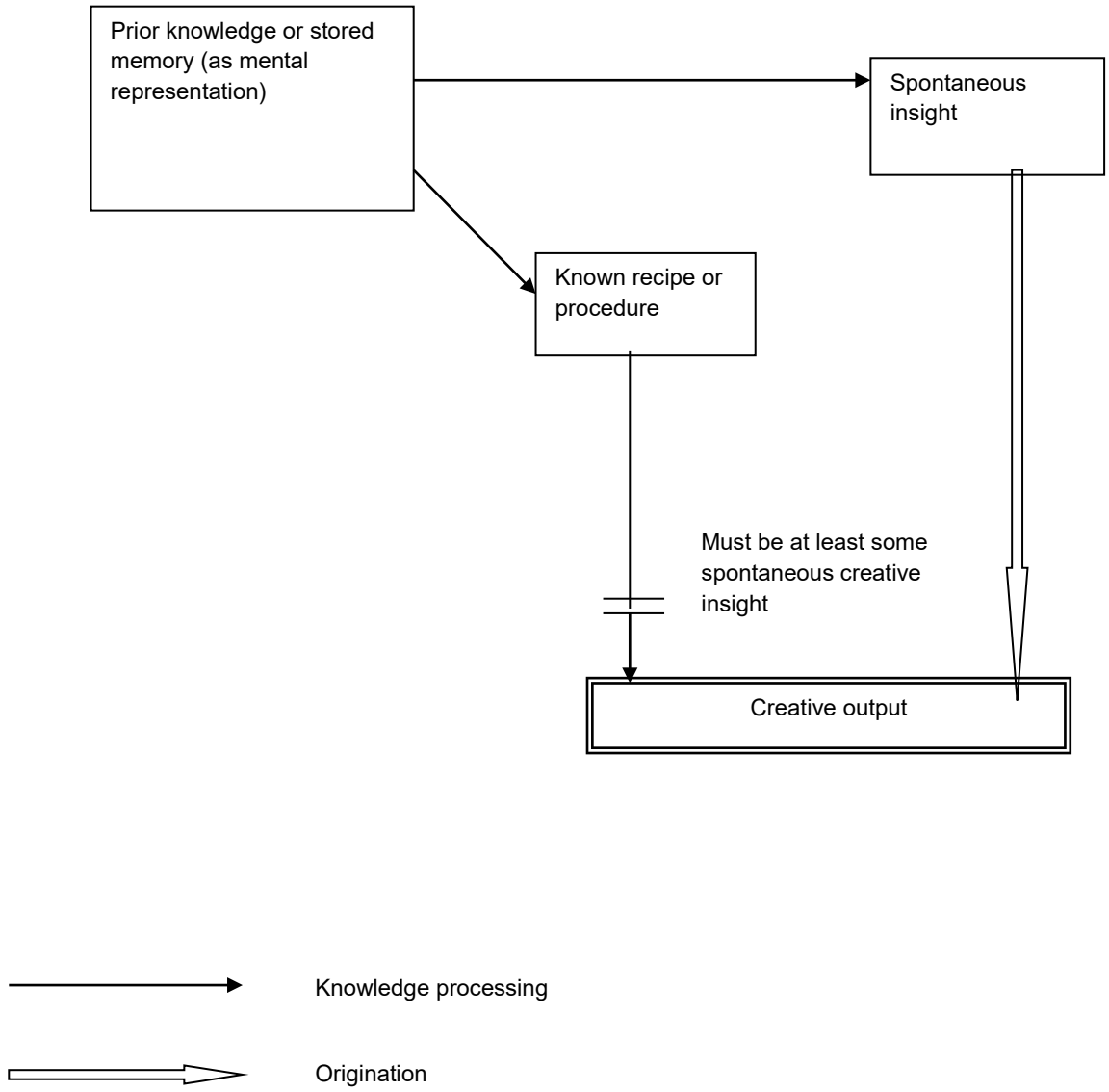
Kronfeldner's originality and spontaneity conditions for creativity, and their associated gradual oppositions, specify requirements for relationships between what I would regard as the source inputs to a creative process, and the overall output (or creative product). In my diagrams following, I have sketched the implications of these conditions for any creative process.

Fig. 5.4: Copying, learning, and creative output



As Kronfeldner points out (2011, p.37) a representation formed in the mind of any feature of the outside world (from sense data, or experience) is a kind of copy, but in another ontological realm, whereas a representation of anything communicated socially has the same ontology as the original. But in each case, Kronfeldner's originality condition demands that creative output is to some extent independent of prior knowledge, which I have interpreted as requiring a novel combination, with additional input from elsewhere within the individual's mental content.

Fig. 5.5: Spontaneity and routine production.



To be classed as creative, a novel idea cannot be drawn purely or directly from prior knowledge, otherwise it would not be new. Something fresh must be added, exactly as Kronfeldner says. Here, I am conjecturing that the spontaneous insight corresponds to the formation of a new combination within the creative agent's personal possibility space, or rather to its recognition as being of potential relevance relative to a known set of required properties.

On my own conjecture, learning and prior knowledge presumably supply the internal component of the universe of source elements available to the creative person, the external component being those parts of the outside world which trigger or key serendipitous discovery. I am suggesting that any specific creative thought only becomes possible given the presence of the required components – even if the novelty generated in a creative process is not itself a property of the constituent elements in their uncombined state, rendering those components necessary, but not sufficient, for the creative idea to manifest.

The outcome of the above discussions on Kronfeldner's originality and spontaneity conditions will be carried forward to inform the modelling exercise.

The Poincaré/Wallas four stages

As said above, the four stages might map approximately onto a Darwinian scheme. Preparation provides the necessary prior information input, thus setting up the internal possibility space, and also for the formulation of the requirement. Incubation and inspiration would then correspond to the variation aspect. Elaboration and evaluation would seem to incorporate an element of selection. It might be that a moment of insight cannot but seem mysterious, given that it reflects the successful result of a search. It may well appear very much as if it comes from nowhere or nothing, as the search outputs something with an element of novelty, which must necessarily be previously unknown.

Varieties of variation relative to creativity and the generation of novelty

Blindness, spontaneity, and the *Meno* paradox all bring in the question of the relationship of a novel, creative idea to previously acquired knowledge. The outcome cannot be known (or be accessible) beforehand. Thus in my own conjectural scheme, the prior knowledge informs and directs the search and the subsequent selection but is insufficient to generate the novel solution by itself.

Random variation with equiprobable outcomes would be ruled out by the sheer vastness of the scope of combinatorial possibilities, as evidenced by the *Universal Library* problem, notwithstanding that the range of such possible options or possibilities might be limited to those accessible to the mind of the creative individual, that is the internal, or personal, possibility space.

Kronfeldner's conclusion that an analogy based upon unjustified variation would not do much to advance understanding can be accepted, without adversely affecting the efficacy for current purposes, because, as said earlier, this (lesser) degree of fit to a Darwinian analogy may well be adequate for current purposes, given that a strict parallel (incorporating undirected variation) is not being proposed. Indeed, if blind variation were to be taken in the neo-Darwinian sense, as undirected, then this would run counter to the perception of creativity as being in some way steered by the human agent, with the benefit of prior knowledge, as Kronfeldner points out.

Prior knowledge and the examples of Kekulé and Fleming

In Kekulé's case, the factors influencing variation could not have been entirely separate from the selective factors, as Kronfeldner says. On my conjecture, the search would be focused on properties, but would rely on structured prior information. The coupled selective factors presumably did not incorporate any reference to snakes, although their crucial property, the facility to form closed curves, features in Kekulé's solution. This situation therefore highlights points of difference from organic evolution, because the straightforward analogy calls for complete decoupling, and biochemical replicators are not equipped to direct a search.

It might be noted here, that insofar as the individual's possibility space is grounded on their already acquired knowledge, then it will be specific to the individual concerned, because of their unique prior experience. This in itself might constitute a source of variation (between individuals) and perhaps even a semi-random one at that.

In my terms, Kekulé's creative idea was sourced internally, as opposed to Fleming's serendipitous observation, which is less obviously coupled to selection. Recalling the earlier discussion on the formulation of a requirement, it might be reiterated at this point that although Fleming was not currently engaged on the problem, he had sufficient prior knowledge to understand the need for an antibacterial agent. It was this awareness that enabled him to recognise the solution when he encountered it, as Kronfeldner points out (2011. p.57).

As said above, the presence of coupling may well prove helpful for modelling internal variability. The proposed requirement, search, and recognition sequence will introduce a significant component of directedness.

Unconscious variation

Poincaré thought that new ideas were born as combinations formed in the unconscious mind, from elements which had been to an extent preselected through conscious preparatory work. These elements are then set free to form combinations:

> The mobilized atoms are therefore not any atoms whatsoever; they are those from which we might reasonably expect the desired solution. Then the mobilized atoms undergo impacts which make them enter into combinations among themselves or with other atoms at rest which they struck against in their course. (Poincaré, 1908, p87).

Promising combinations are then selected for elevation to awareness according to mathematical elegance:

> Among the great numbers of combinations blindly formed by the subliminal self, almost all are without interest and without utility; but just for that reason they are also without effect upon the esthetic sensibility. Consciousness will never know them; only certain ones are harmonious, and, consequently, at once useful and beautiful. They will be capable of touching this special sensibility of the geometer of which I have just spoken, and which, once aroused, will call our attention to them, and thus give them occasion to become conscious. (Poincaré, 1908, p86]).

This approach as it stands is open to the *Universal Library* objection due to the potential numbers of combinations. As Kronfeldner (2011, p.65) points out, Poincaré

understands this and introduces a proportion of direction via the preselection as above (but the degree of preselection would presumably have to be very high indeed to escape from the combinatorial problem entirely, see chapter two).

The Darwinian analogy in comparison with ordinary mental processes

Of particular interest for current purposes is Kronfeldner's reference to Jonathan Schooler and Sonya Dougal's (1999: 352) observation:

> that spreading activation, for instance, is "by no means random in nature". On the contrary, it depends not only on the initially activated sample of items, but also on the "underlying structure of an individual's knowledge representation".[…] (2011, p.66, reference to note omitted).

This bears a superficial resemblance to a query or search of structured information. The proposed model of internal variation and selection, by directed querying and searching, based on a known requirement, would presumably count as an example of Kronfeldner's ordinary-process view.

Directed variation and the proposed model

I would suggest here that the basic Darwinian analogy becomes non-trivial when appropriate modifications (the requirement, search, recognition sequence) are brought into play, as discussed at various points in the second part of the chapter. To this I would add that guided variation is potentially more effective than the undirected version characteristic of evolution by natural selection, which is as might be expected when the human context of creativity is taken into account, given the mental faculties which can be brought to bear, and which are unavailable to evolution in a biological setting.

5.5 Conclusion to chapter five

It is now possible to draw together and summarise the results of the above discussions, focusing on the outcome of the feasibility exercise, and also on those findings which have a bearing on the construction of a model.

5.5.1 Methodological issues

As said, modelling will be based primarily on boxological methods, of necessity restricted to a high level of aggregation, and a sub-Fodorian modular structure will be taken as a framework. The benchmark criteria for success will be as set out earlier in this chapter

Modelling will draw upon the results of the investigation of the proposed Darwinian analogy, incorporating appropriate amendments and refinements as detailed below. The extended list of stages (as described under 'Stepping through a creative process' above), has been reviewed at appropriate points during the discussion of the proposed Darwinian analogy, and no major revision seems to be called for as a result. Therefore this scheme seems suitable to be taken as a departure-point for modelling.

5.5.2 The Darwinian analogy

The evaluation of the proposed use of Darwinian evolution as an analogue for creative process is now complete. The analogy has been subjected to careful examination by Kronfeldner, as described above. Kronfeldner refers to a strict Darwinian analogy throughout, and therefore her arguments will not necessarily apply, or at least not in the same way, to the kind of modified analogy being proposed. But they draw attention to – and sharpen – the points of difference between creativity and evolution by natural selection and therefore provide indications as to what a modelled creative process has to accomplish. The following are among the points that can be extracted from the discussion for particular attention in chapter six.

It now seems clear that there are significant divergences between Darwinian evolution and creativity, which appear to be relevant to the processes concerned. They seem likely to frustrate any attempt to understand and model a creative process by reference to a straightforward analogy alone. There do remain some basic similarities nevertheless, Including variation and selection, the origination of novelty, and the

generation of added value (or fitness), so I will not need to abandon the use of an analogy altogether. In my own view, the discrepancies reinforce the case in favour of taking a modified analogy as a pattern. The differences identified will need to be carried forward to the modelling exercise. Kronfeldner's analysis has highlighted, and helped to refine, the issues that will need to be taken into account.

Creativity involves at least two stages of selection. Innovations proposed by individuals are subject to selection at the cultural level, but they are first pre-selected by the individual prior to being put out for approval by the wider culture. At the cognitive level, the Darwinian analogy does seem to fit with creative process in the weak sense of requiring variation and selection. In that form, it has survived the above investigation, but it is of little assistance in explaining creativity as it stands, as Kronfeldner points out.

Acquisition of prior knowledge is essential for creative thought, although to generate novelty, any candidate creative idea must add something over and above the sum total of such information. Prior knowledge can be used to frame a requirement, to define the criteria a creative product needs to meet: this is the main plank of the method to be put forward as avoiding the *Meno* trap. Once established, the requirement can be used to direct a search. This is a point of difference from Darwinian evolution as it will involve coupling between selective and variational factors. The search may be directed externally, at aspects of the individual's environment, or internally, at stored information in memory, for which to be successful the information store will need to be appropriately structured to facilitate such access. Search outcomes will need to be matched against the requirement.

5.5.3 Sources of novelty

If *ex nihilo* creation of ideational material is ruled out, there are two possible sources: internal to the individual and also in the external environment.

In the case of purely internal sourcing, directed variation (as in a requirement-directed search) can be fed into the model with reasonable confidence. There will be resultant incorporation of coupling between selective and variational factors, so that the strict Darwinian pattern will not be used for internal searches.

The scope of internal search is limited by the individual's personal possibility space (the total of all possible combinations of elements of their stored information). This space is a subset of the universe of possibility, as discussed in chapter two. Variation in the range and boundaries of the space, together with differences in search strategies, would lead to corresponding variation in output between individuals, even without allowing for any degree of undirected or random variability. The external environment may be subjected to directed search (as in laboratory trials, for example). There is also the possibility of serendipitous observation, which provides an undirected, random, or pseudo-random source of input.

If a model is based on the conclusions for this chapter, therefore, it will allow for directed internal and external search, with the option of supplementary input via external serendipitous clues. This will provide the variation component, and will be followed by internal pre-selection.

Chapter Six: *Modelling a Creative Process*

6.1 Introduction

I now have in place all the essential components needed to begin the modelling exercise.

From chapter three, I can draw the modular framework against which the model can be structured. The strict Fodorian criteria for modules will need to be relaxed on a couple of dimensions to allow a degree of disencapsulation, so that data held outside the module can be made accessible to its internal processing.

In chapter two, I discussed the Wallas four-stage model of creativity, and I began to develop it into an extended sequence, to include the formulation of a requirement, and the use of this to drive a search and to recognise potentially successful novel thought. I firmed up the sequence in chapter five, by adding a further step to allow for cultural preservation, and mapping the stages onto an equivalent series derived from biological evolution, giving seven stages overall. This extended sequence can be used to form the spine of the modelled process.

I am proposing to build a modified Darwinian process, with variation and selection phases as in organic evolution. In chapter five, I noted the apparently directed character of the variation aspect of creativity (as opposed to the undirected variation of the strict Darwinian view). A number of discrepancies became apparent with respect to the Darwinian scheme that appeared to reflect the human setting, with its faculties of design, awareness, and analysis, which are unavailable to biological evolution. I observed that one major difference was the ability to accumulate prior knowledge in the area of the problem, and to utilise that knowledge to understand the nature of the

current environment and the challenges it presents, and then to formulate a requirement corresponding to the properties desirable in a creative product or output.

I pointed to hints that prior knowledge might be stored in a structured form, which will include links on the basis of properties, so that the requirement could be used to direct a search of stored material.

A search may be targeted internally (at stored prior information) or externally (at information available without, such as in nature, culture, records, or from testimony) and/or also at serendipitous observation. On this view, therefore, these locations would hold the source input to a creative process. I have assumed that, in the absence of any *ex nihilo* contribution, all creative output must be sourced from existing ideational elements, though in novel combinations; emergent properties absent from the source material may arise as a result. The range of potential novel combinations will be limited by the individual's personal possibility space – the set of combinations which can be constructed from all ideational elements available to them, whether externally or internally derived.

The central concern of the project is the question of the origination of the novelty – how new ideas might come to be generated. This corresponds to the variation aspect of a (biological) evolutionary process. I am therefore putting correspondingly less emphasis on the selective aspect, though creative novelty has to add value, and selection is thus an integral part of any creative process. I plan to model this by adopting (or adapting) an arrangement similar to the possible worlds box or supposition mechanisms put forward by Nichols and Stich, and Carruthers.

I will therefore start the modelling exercise with the following building blocks: a modular framework; a staged sequence; requirement (for properties), search, and recognition steps; a weakly Darwinian (guided variation) mechanism; pre-existing source material; and selection via possible worlds simulation.

6.1.1 Coverage

This chapter focuses on the central thought experiment of the study: the construction, as far as possible and practicable, of a high-level model for a process which might be expected to produce creative output. The process will need to be capable of generating novelty and of assessing and selecting according to value. The chapter is divided into sections as follows: model foundations (structure and processing); preliminary boxological sketches; model construction; and counterarguments. Counterarguments from Carruthers and Boden are explored later in this chapter; a more comprehensive evaluation of the model follows in chapter seven.

6.2 Model foundations

The guiding principles for the modelling exercise can now be summarised. They take into account the outcome of the feasibility assessment, together with those conclusions from earlier chapters which have a bearing on the modelling exercise.

1. It is reasonable to attempt the construction of a model, no objection having been identified which would rule this out.

2. The model is to be put together as a high-level boxology, assuming a sub-Fodorian modular structure.

3. The staging of the process will follow the extended Wallas scheme, as developed in earlier chapters.

4. The process will be assumed to operate on essentially Darwinian lines, subject to certain caveats, as below.

The Darwinian analogy survived the examination in the previous chapter. Counterarguments from Kronfeldner can be accepted without compromising the approach if the analogy is taken as a basis for modelling the process (although not if taken in itself as an explanation of creativity). The strong, strictly Darwinian,

interpretation requiring undirected variation (decoupled from selection) is only suitable, if at all, for serendipitous input where this is a random element. Otherwise, a directed search is proposed.

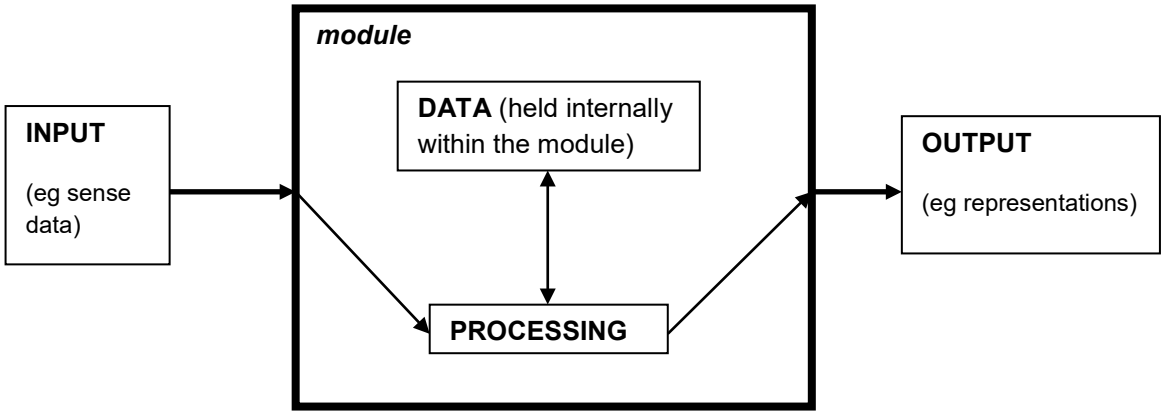
6.3 Preliminary sketches for modelling

6.3.1 Modularity

The underlying structures will follow the modular, or massively modular, pattern (although they will initially be put together as process boxes, for ease of construction).

The context of a classic Fodorian input module was presented in diagrammatic form in chapter three but for modelling purposes it is necessary to look more closely at its internal structure and processing, which I have sketched in the diagram below.

Fig. 6.1: Classic Fodorian input module (encapsulated, domain-specific)

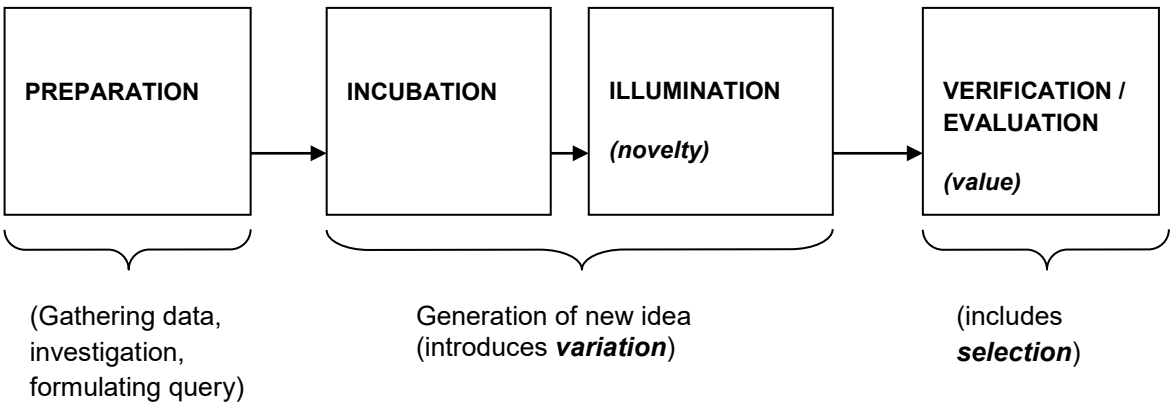


The module is domain-specific in that it is restricted to a single mode of sensory input (vision, for example). It is encapsulated in only having access to data held within the module itself. In the Fodorian architecture, creative processing would presumably be located within the central systems, which Fodor regards as nonmodular. Under massive modularity, a central systems process may be expected to require one or more sub-Fodorian modules.

6.3.2 Staging

Wallas has the four stages illustrated below (with correspondences to variation, selection, novelty, and value).

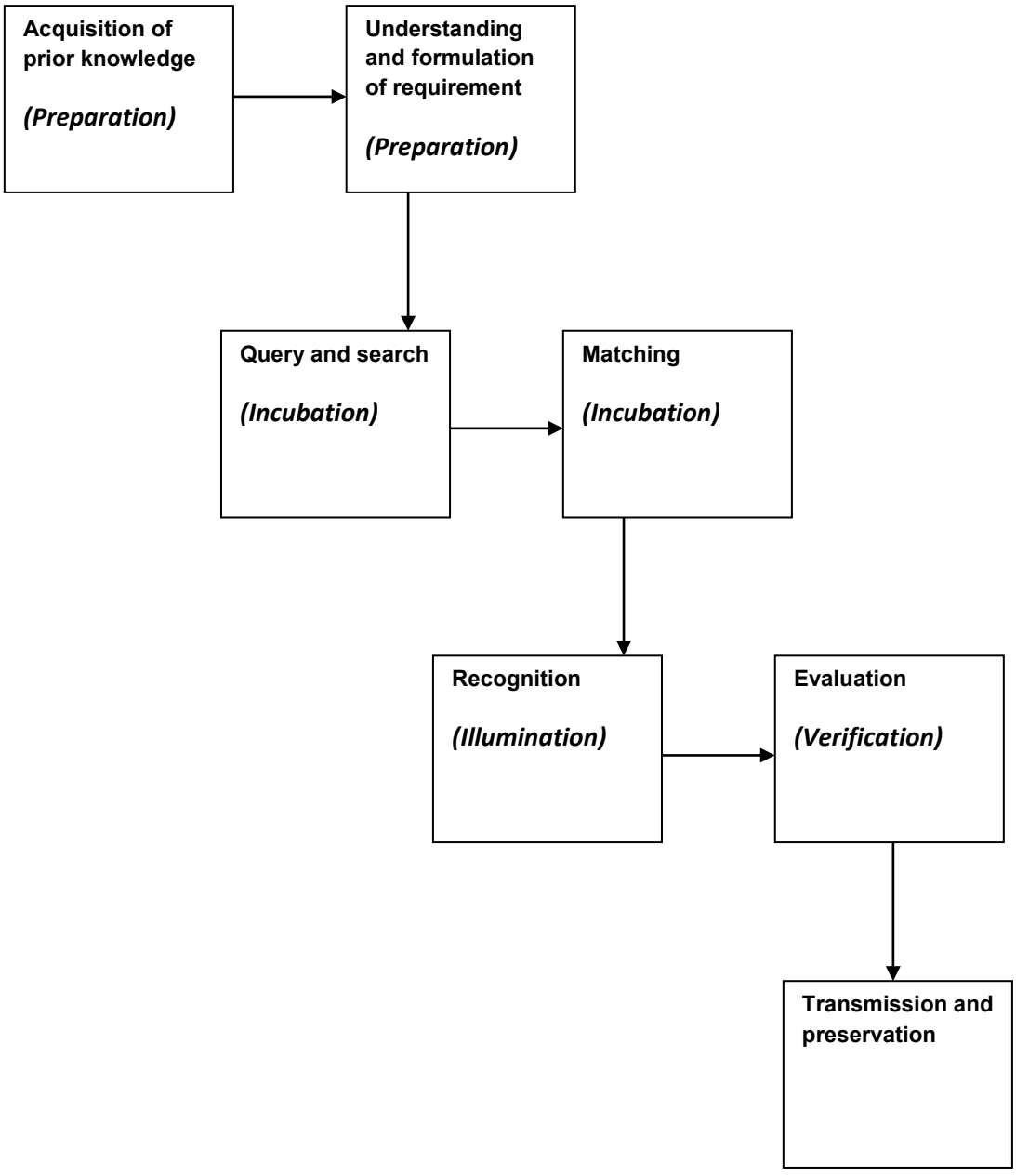
Fig. 6.2: Wallas stages (annotated)



Here, the preparation stage carries the accumulation of prior knowledge. This will be insufficient as it stands to generate novel thought, which makes an appearance in the second and third stages. The final stage, verification and evaluation, is the testing of the new idea, a selection process to check that it adds sufficient value to count as creative and so to be considered for implementation or wider transmission.

The proposed extended staging – from chapters two and five – runs as follows, with approximate mapping onto Wallas (the Wallas equivalents are indicated here in brackets):

Fig. 6.3: Extended staging

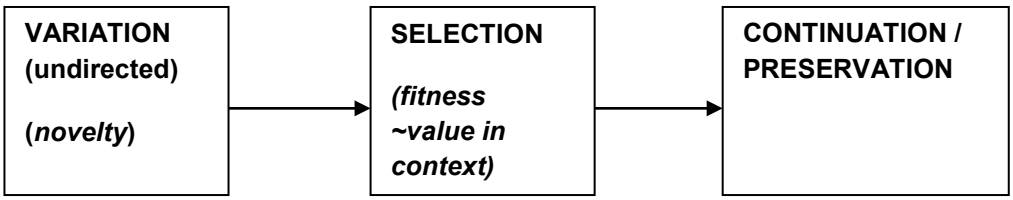


The extended version of the staging brings in the analysis of the prior knowledge to inform an understanding of the properties desirable in a creative product. It is this requirement that is used to drive the search, which is here taken to occupy the incubation period. A successful matching of the search output against the requirement amounts to recognition, a potential match (or partial match) of the required set of properties has been identified. This is then checked and evaluated in the verification stage, and if appropriate considered for transmission or implementation.

6.3.3 Darwinian evolution

A strict Darwinian process has variation and selection stages, and the variation is undirected; the relative frequencies of variants are not influenced by selective factors. I have represented this arrangement in the diagram that follows.

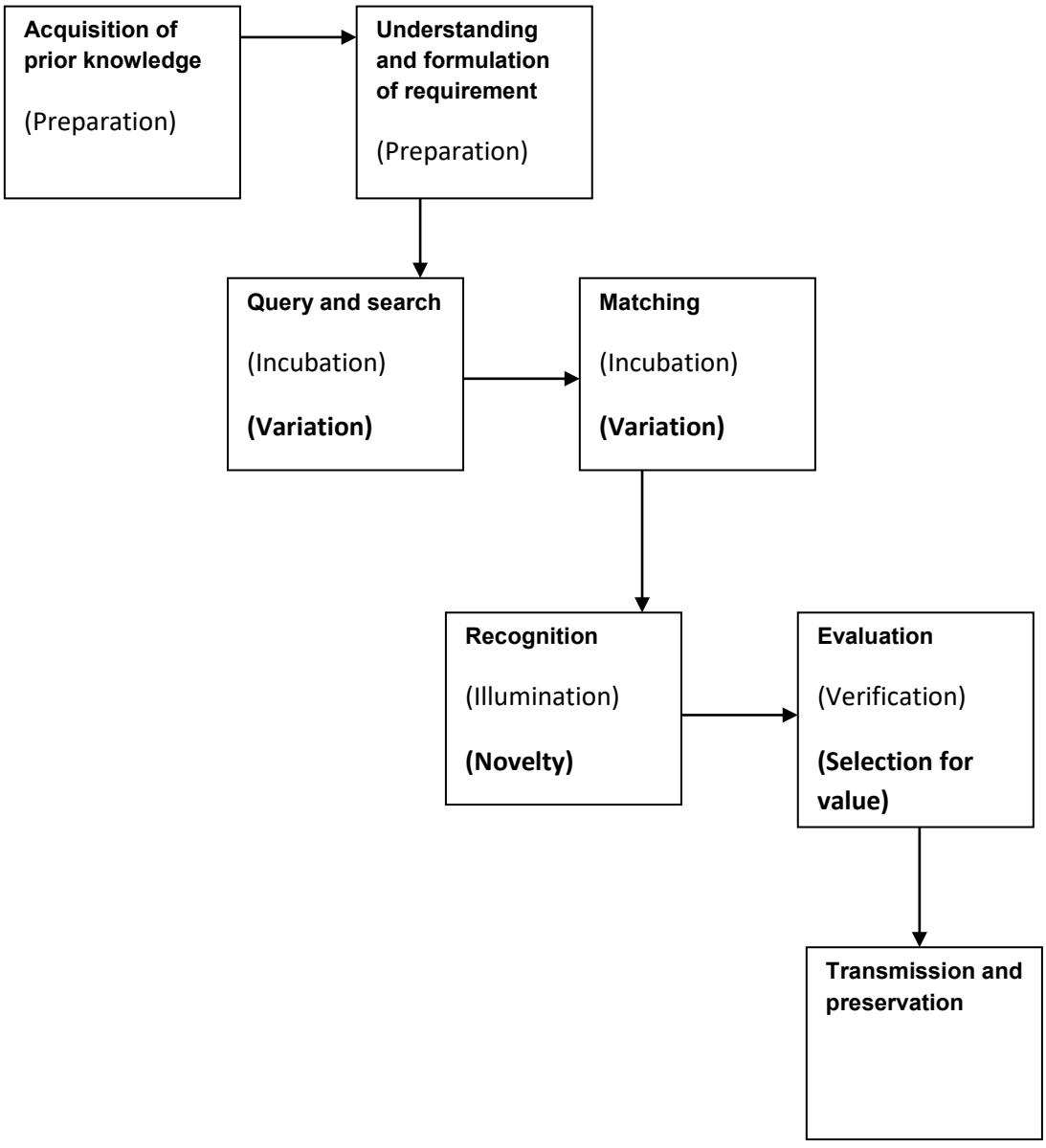
Fig. 6.4: Natural selection (generalised)



The variation step originates novel variants, which are then exposed to the selective factors. Their relative potential fitness is a key determinant of success (see chapter five). If a creative process could be modelled as a straightforward analogue of biological evolution, then I would expect these steps to cross over in much the same form as shown here. But this is not the conclusion arrived at in the previous chapter, which calls for staging along the lines of the extended pattern in the previous diagram, to better represent the situation for creativity as opposed to evolution.

Mapping from the generalised Darwinian process can be added on to the extended staging, but in which case the variation may become wholly or partly directed, as previously discussed:

Fig. 6.5: Extended staging (with mapping to natural selection)



In the combined diagram above, the variation is mediated in the search and matching phases, and is here expected to be directed by the requirement, as opposed to the undirected variation characteristic of organic evolution. Novelty becomes apparent at the recognition stage, and value is established by selection, which is located in the equivalent of Wallas's verification stage (here labelled 'evaluation').

As it stands, the extended staging diagram shows only the overall outline of a process. Detail can be added at a number of points, on the basis of the investigation in earlier chapters; this is particularly so in terms of the data holdings, and of the processing steps which they might be expected to inform.

6.3.4 Prior knowledge and the requirement

At the commencement of the process as represented in the above diagram, the acquisition of prior knowledge assembles a database of material of relevance to the problem area. This forms a subset of the total information held by the creative agent. The overall internal data holdings define the internal possibility space, the postulated set of all possible combinations of the elements of the stored data. This in turn is a tiny subset of the proposed universe of possibility, discussed in chapter two. Its smallness is of course relative, so that in fact its vast extent precludes any attempt to hold all the possible combinations at the same time, in memory or in any other form. Also, the lack of indexing of any such construction would render searching utterly ineffective, as described for the *Universal Library* in chapter two. Therefore it cannot in any practical sense be drawn upon directly in a creative process. It does, however, represent the theoretical limit of combinations available to an individual for internal search, and hence also of creative thought available to the individual from internal resources alone.

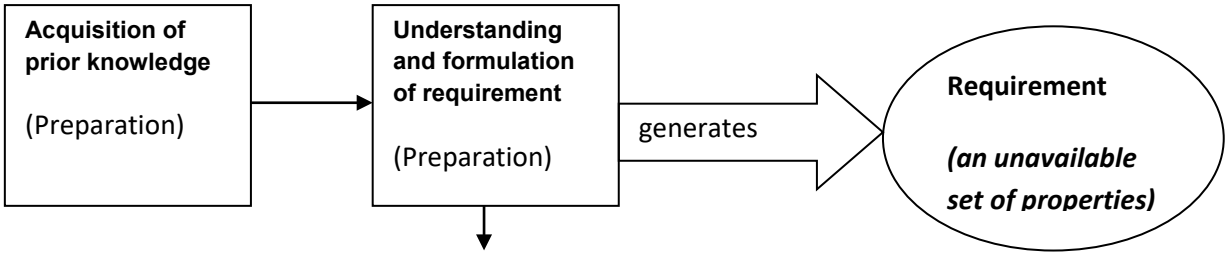
Combinations embodying novel properties and which include elements from without the individual possibility space can therefore only be set up with the assistance of contributions from external search, though it might be possible to add a proviso to this. It might be that the internal possibility space can be enlarged by material emergent

from previous creative thinking. For example, having brought together snakes and carbon chains to make the benzene ring, Kekulé might perhaps have been in a position to consider the possibility of combinations of chains with rings, such as found in toluene (whether or not he actually did so). Such a development might be considered to have expanded his internal possibility space without recourse to external observation. In other words, it might be that the internal possibility space can be extended by means of creative thinking, due to emergent novel possibilities.

The target of the prior knowledge stage consists of material of more-or-less direct, or background, relevance to the problem area. Following acquisition, this defines a sector of the individual's possibility space of particularly promising potential for a directed search, although of course novelty might be derived from elements in entirely different domains, as for Kekulé's daydream of snakes, which do not bear any close links to organic molecular structure. The individual has no internal route of access to material beyond the boundary of the personal possibility space other than by emergence as described above, unless provision is to made for *ex nihilo* creation of ideational material (this would not on the face of it seem any less problematic than *ex nihilo* creation of material entities, if intervention is to be ruled out as a source of novelty).

The creative agent's prior knowledge, while not on its own providing a comprehensive, ready-made target for searching, does allow for the formulation of a requirement. This is just an acknowledgement of gaps, of sets of properties which would have utility in the agent's internal model of the external world, but which are not known to exist in the appropriate form and combination needed for the unfulfilled purpose. The requirement can therefore be added to the diagram as an information holding arising from the acquisition of prior knowledge:

Fig. 6.6: Generation of the requirement

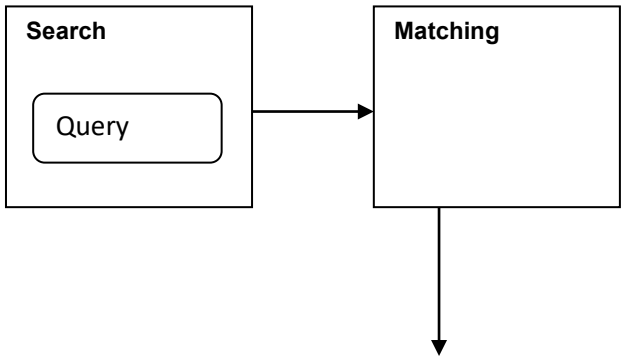


Here, I have added in the requirement as a data (as opposed to process) box; the requirement will be drawn on in later stages, as follows.

6.3.5 Search and query

The acknowledged requirement can then be used to direct a search. The goal of the search is to identify a combination of properties sufficiently similar to those in the requirement to fulfil the implied purpose. Thus the requirement is a set of properties: $\{P_1\ P_2 \ldots P_n\}$ and the search function sets up a query to find items with that set of properties, or some of its members.

Fig. 6.7: Query and search



For example, in Kekulé's case, his dream produces the idea of snakes forming closed curves. It is not conceivable that it connected this with his current work on the structure of benzene by any randomising mechanism, because of the vast extent of the internal possibility space. But, he knew from earlier work that carbon atoms formed chains, or strings, and a search for items with string-like properties would be much more focused and efficient. It implies a structure facilitating such a search; the internal store of information cannot be unstructured, like the *Universal Library*, or specified properties would not be retrievable. It is this kind of direction that can allow the creative process to improve on the efficiency of variation and selection over and above their manifestation in organic evolution, and it is in this way that the divergences from a full analogy may be as significant as the correspondences.

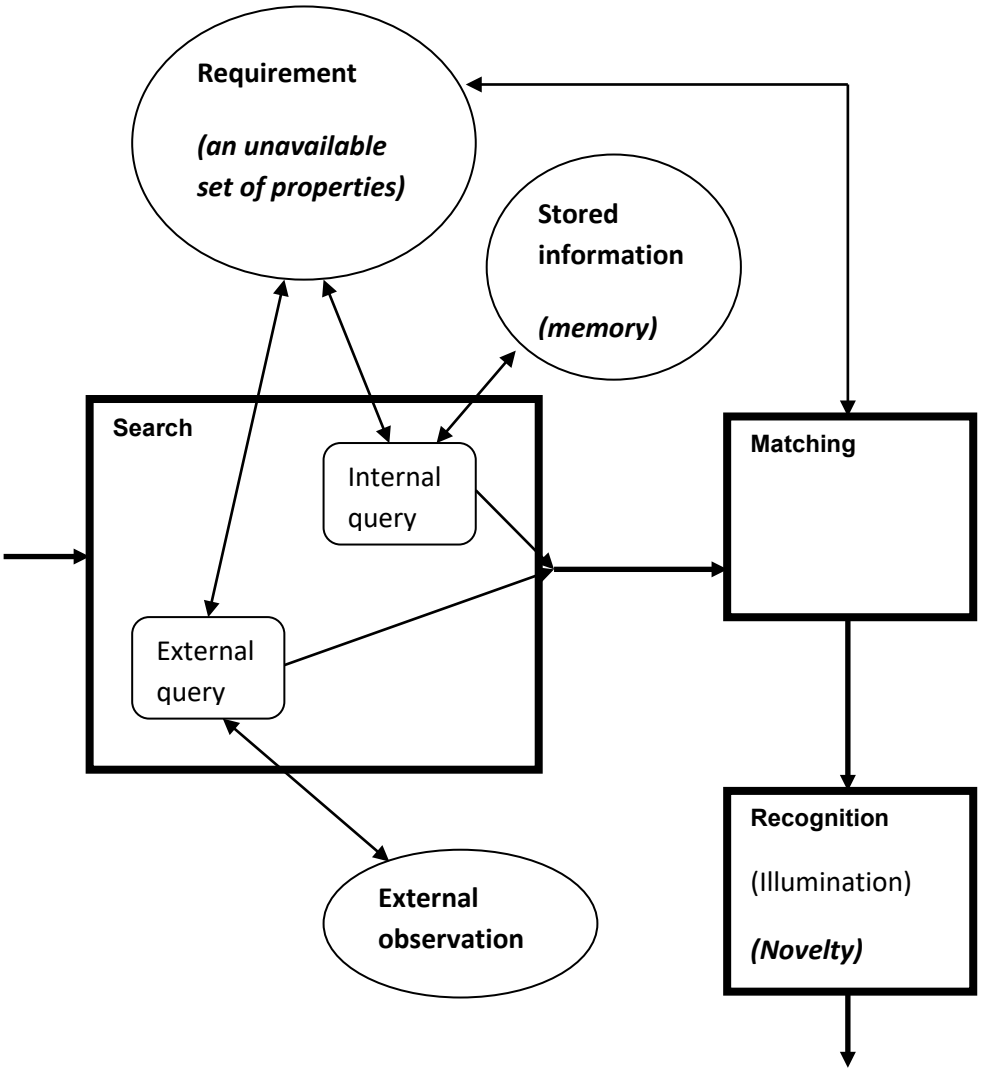
As said, the scope of internal search is restricted to the possibility space – to explore beyond its boundaries requires external input, or perhaps emergent novel ideas from earlier creative thinking. Here it is suggested that external as well as internal search might be directed by the requirement. In Fodorian terms, it would be necessary to search the representations from sensory input modules for the set of properties defined by the requirement. This is presumably similar to searches by other animal species for the presence of desirable items in their surroundings, for example prey, or potential

mating partners. It might be anticipated that such searches would also be requirement-driven: features of the environment would be matched against criteria (sets of properties). For example, small (possibly erratically) moving objects might indicate insect prey, or a perfectly formed and displayed tail might signal for a genetically fit male peacock. Thus Fleming's serendipitous discovery of penicillin might also be thought of as having been driven by a set of criteria – or a requirement – for antibacterial properties.

Serendipitous contributions to creative process inevitably bring in a random element, so that externally sourced creative novelty approximates more closely to the strict Darwinian analogy (see chapter five). As previously said, the likelihood of encountering promising serendipitous material might be increased by choice and control of environmental factors, so that Fleming's being situated in a laboratory setting increased his chances of encountering the action of penicillin mould.

The search box can now be expanded:

Fig. 6.8: The search box



The diagram provides for the search box to run internal and external queries. It has to have access to three types of external data. All queries draw on the requirement, the set of properties established through the analysis of accumulated prior knowledge. Queries also need to access target data for searching, which may be stored prior information or externally derived data, as appropriate. Thus the search box cannot be fully encapsulated, although the data needed to run the query functions themselves can be held within the box.

6.3.6 Matching and recognition

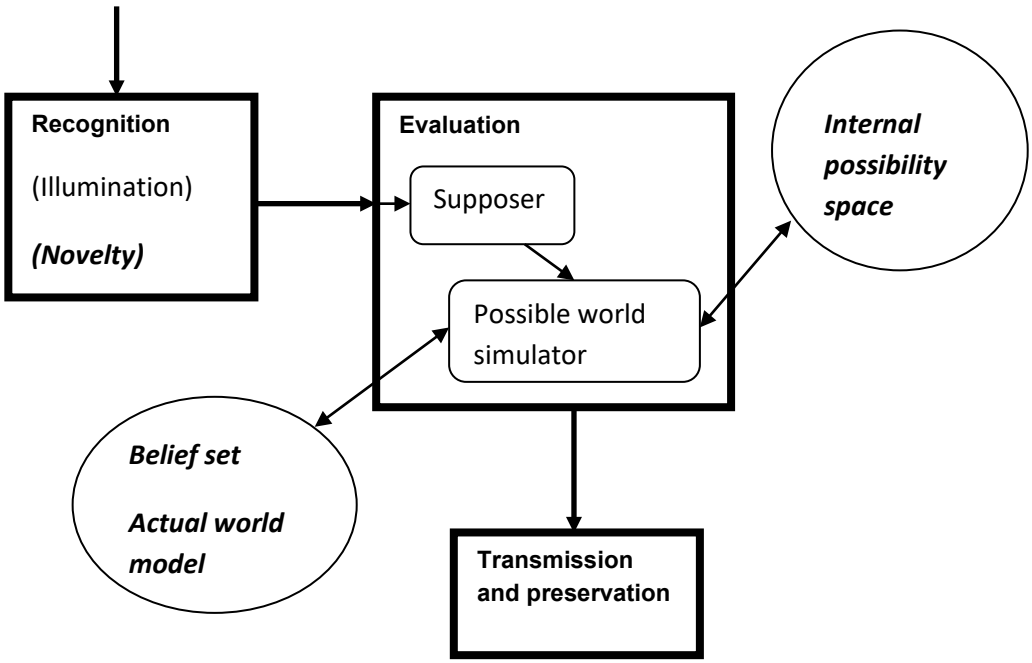
The novel combinations of properties resulting from searches will need to be matched against the requirement, for goodness of fit. A successful match can then be recognised as a candidate novel idea, corresponding to Wallas's illumination stage: the first appearance of the new.

I have included matching and recognition boxes in the above diagram. The matching process needs to draw on the requirement, in order to make comparisons with the output of the search box. A relatively successful match triggers output to the recognition box. Receipt of this as input to recognition might correspond to, or signal, a moment of insight.

6.3.7 Evaluation

Following on from recognition, there is evaluation, the selective part of the process. This stage is less obscure, or mysterious, than the origin of novel ideas, but is nevertheless an essential part of the process, needed for establishing value. Carruthers (2002, 2006) refers to supposition: evaluation can be carried out by supposing that the new idea were to be adopted, and setting up a possible world model within the individual's possibility space, in which the supposed situation obtains.

Fig. 6.9: Recognition and evaluation



In this diagram, output from the recognition box, indicating a potentially valuable match, is sent on to an evaluation box. Evaluation is carried out in a mental trial, through the supposer and possible world simulation operations. The possible world simulation needs external access to the current set of beliefs about the actual world (the actual world model), and also to the internal possibility space, in which the possible world dependent upon the supposition is constructed. The function of the possible world

simulator is much the same as the 'possible worlds box' proposed by Nichols and Stich (2003), from which it is derived.

The use of supposition at this point in the model also draws on possible worlds theories (of, for example, Lewis (1986), and Alvin Plantinga (1976)). It is taken as an initial assumption that *to believe that P* is to believe that state of affairs P obtains (or did once obtain) in the actual world. The set of such beliefs held by an individual at any one time then constitutes that individual's current model of the world. Thus, to believe that P sets up an internal model of the external world. So, if the individual's current model of the world consists of beliefs B_1 to B_n, then:

$$\{B_1\ B_2 \ldots B_n\} = [\textit{world model}]$$

(This ignores numerous complicating factors, such as the linkages and relations between particular beliefs).

Entertaining a supposition effectively sets up a temporary model of a possible world, by adding the supposition to the current belief set (or substituting the supposition for a currently held belief).

$$\begin{bmatrix} actual\ world\ model \\ (belief\ set) \end{bmatrix} + [supposition] \rightarrow \begin{bmatrix} possible\ world\ model \\ (belief\ set\ incorporating \\ the\ supposition) \end{bmatrix}$$

The supposition is now equivalent to a true belief, but within the new possible world.

The resultant possible world model can be used to evaluate the supposition, by mentally exploring its consequences, in effect running the possible world model to find out whether the supposition leads to added value.

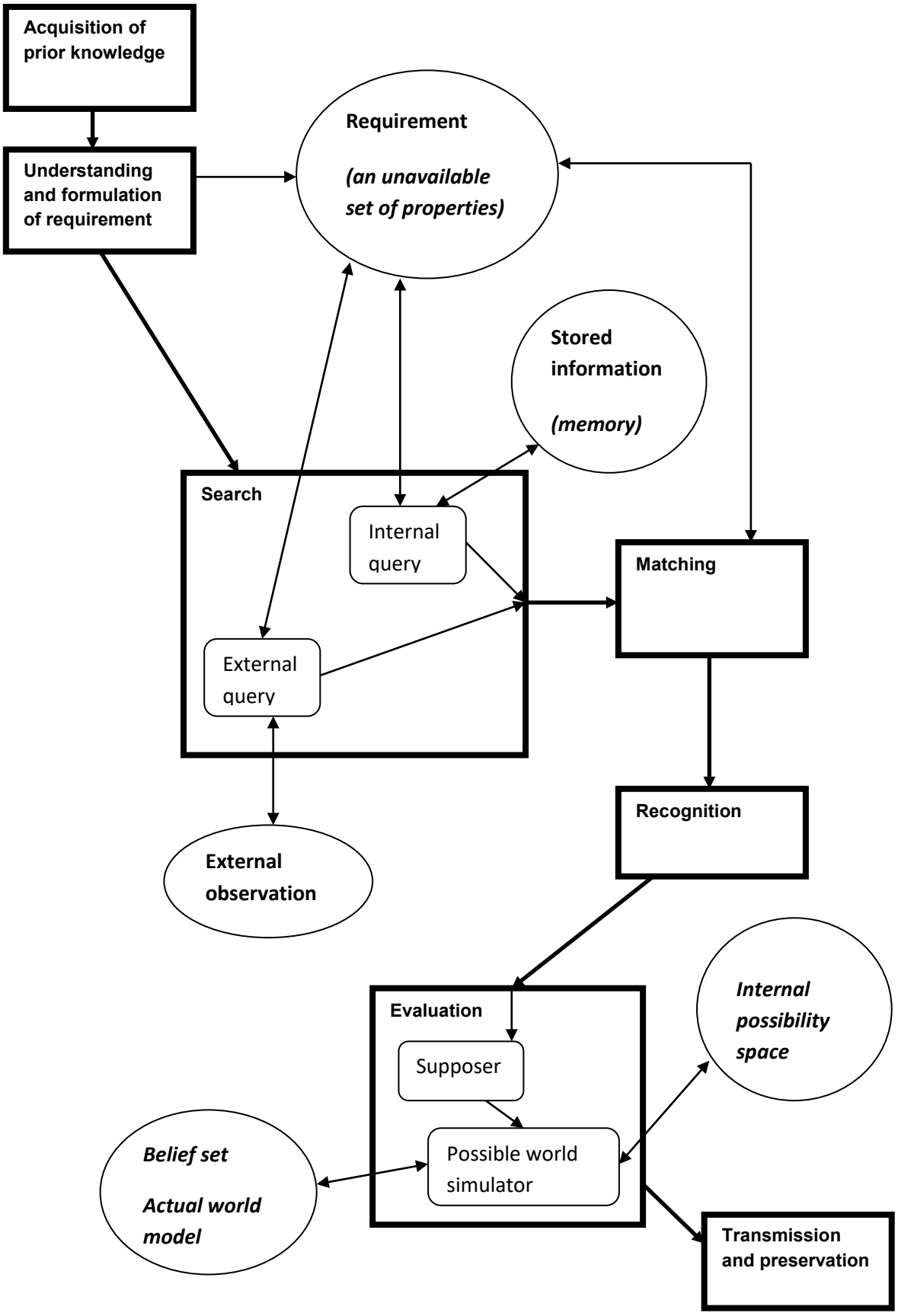
$$\begin{bmatrix} possible\ world\ model \\ (belief\ set\ incorporating \\ the\ supposition) \end{bmatrix} + \begin{bmatrix} inference \\ (from\ run\ of \\ possible\ world\ model) \end{bmatrix} \rightarrow \begin{bmatrix} evaluation \\ of\ likely \\ outcome\ of \\ supposition \end{bmatrix}$$

If the evaluation is favourable, the result is output for consideration for implementation in, or transmission to, the outside world.

6.4 Constructing the complete high-level model

The component parts of the creative process model can now be reassembled into a complete diagram covering the entire process.

Fig. 6.10: Complete process model



The main flow through the modelled process conforms to the extended staging system proposed earlier on. Access is required to external data for searching, and to draw on the internal belief set, or model of the actual world. Additional data holdings are developed and used during the process, corresponding to the requirement and the possible world model.

The pathway through the model is of course greatly simplified; in practice multiple iterations and false starts can be anticipated as the process unfolds.

6.5 Modularisation

In terms of modularity, the diagram already reflects a quasi-modular approach because the process boxes operate as if they were relatively self-contained proto-modules. Some, however, approach more closely to a simple modular structure than do others; some accept input from more than one domain of thought, and/or they draw on data from outside the process box. So where should the boundaries of the proposed sub-Fodorian modules be drawn?

The acquisition of prior knowledge is clearly a complex process, involving the operation of a number of functions, including (on a Fodorian scheme) sensory and language input modules and holistic central systems processing for belief fixation. On a massively modular view, many more modules may be involved at central systems level. The prior information, often comprising not only that having a direct bearing on the problem, but also more widely-spread background, will most likely cross several domain boundaries.

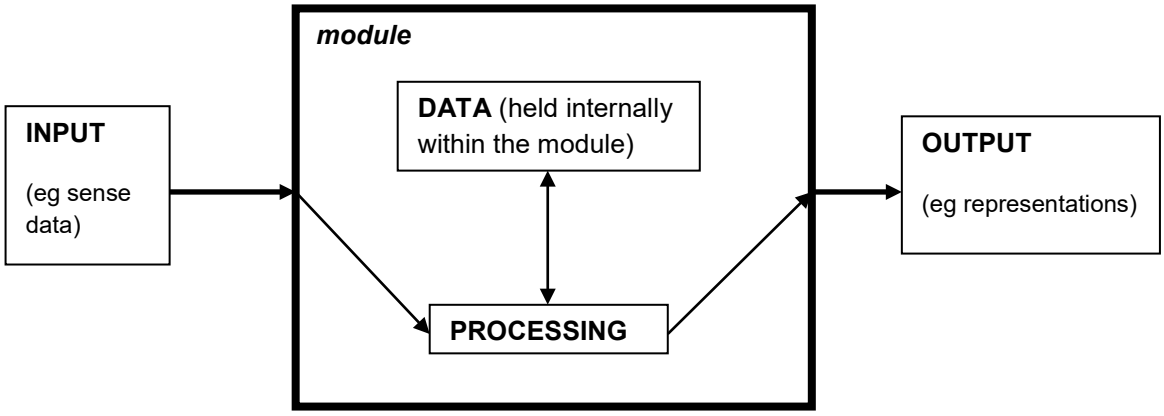
The formulation of the requirement will take place in the central systems and (along with all the more downstream parts of the process) will therefore occur beyond the reach of any of the basic Fodorian modular structure. Probably, then, both the acquisition of prior knowledge and the formulation of the requirement which derives from that prior knowledge are multi-modular functions. They precede the core variation

and selection part of the creative process, which is driven in the first instance by the requirement.

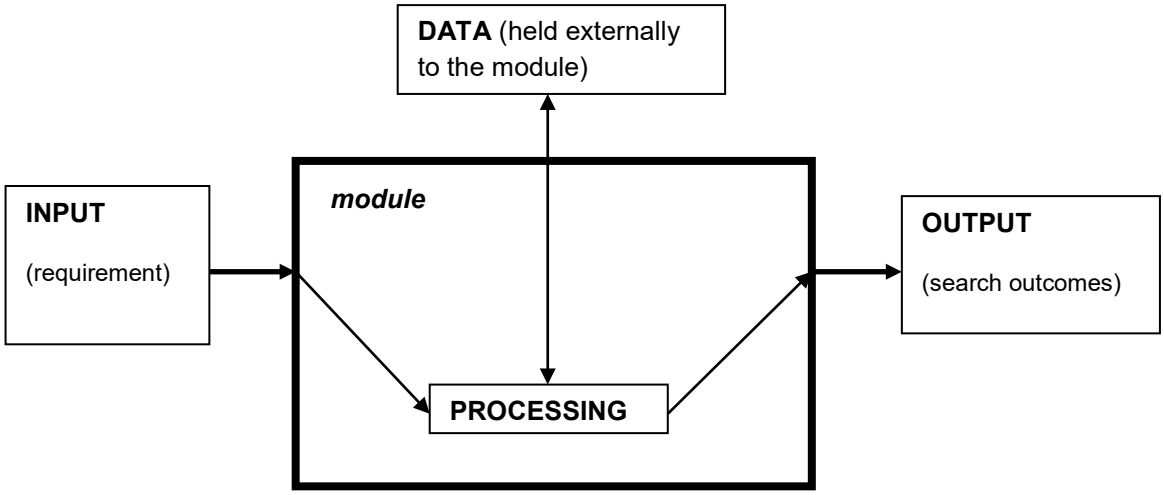
Moving on to the search box, this sub-process – the key to the generation of novelty – particularly reflects modularity. It is a relatively simple, self-contained unit and more closely resembles a single Fodorian module, on the surface at least. It takes the requirement and uses it to construct and drive a query. Data held, Fodorian-style, within the module could be used to form the requirement into a query, and the necessary processing to do this can also be sited within the module. This gives a degree of encapsulation. The requirement itself is built from prior knowledge, outside the search box, and may include information from more than one domain of thought; to this extent the module must be non-Fodorian. However, if the requirement is taken as a single, specific form of input (that is, as similar to a sensory mode or language in the context of Fodorian input modules) then perhaps the resemblance may be taken to be closer than first appearance suggests. The other external data involved, the query targets, is more difficult to reconcile with Fodor. An internal query, as previously said, needs to be free to access any domain of thought, based on properties rather than topic or subject matter. An external query looks for incoming representations that meet the requirement; they might derive from any sensory mode, or from language, or from any combination thereof. Again, this is unFodorian, but it has to do with the nature and function of the module. A search and query module might be expected to differ from an input or output module, precisely because it must draw on external data banks; it has to have an external target area in which to carry out the search. No search can be carried out in the absence of access to databases containing the information sought, so the module as it stands might perhaps be Fodorian in so far is its function permits; Fodorian input and output modules do not need to conduct searches of external data holdings and so the question does not arise in their case.

Fig. 6.11: Fodorian input module and search and query module compared

Fodorian input module (from fig. 5.1)



Search and query module



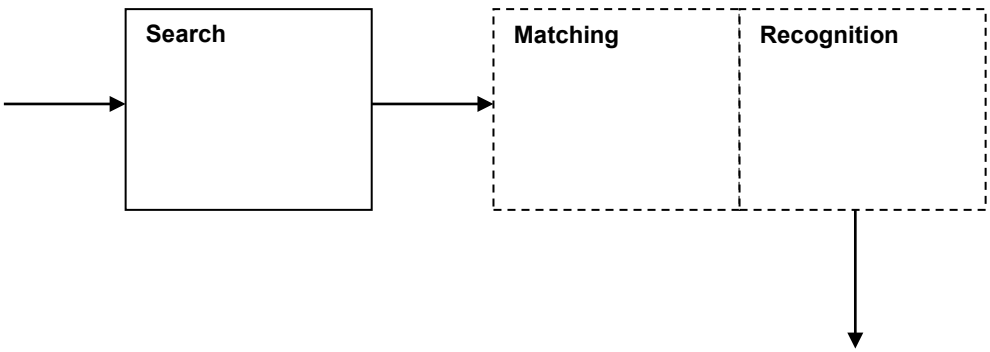
The search box (forming a search and query module) is compared with a Fodorian input module in the above diagrams. The internal and external search functions have been combined for simplicity.

Considered as a module, the search box is necessarily unencapsulated, but is nevertheless self-contained in terms of processing, and in this sense it operates in relative isolation from other mental functions.

The matching box accepts output from the search and query module and checks it against the requirement, for goodness of fit. Input therefore consists of search results, and also the requirement itself, which has been discussed above. Otherwise, the box is self-contained and can use internal data and processing to carry out the matching. Thus it would go some way towards qualifying as a Fodorian, encapsulated, module. It is a separate operation from search and query and is probably best kept outside the boundary of the search module.

Recognition amounts to the signal that a match has been found. It is the first appearance of a novel idea in awareness and corresponds to the eureka moment. The recognition box takes input from matching and outputs a recognition signal, which is passed onwards to the evaluation box. The input and output are relatively specific, and the box can be expected to function with otherwise internal data and processing. It therefore approaches Fodorian encapsulation and domain specificity. Here, the boundary with matching is not so well defined. Matching is the only source of input, so there is a question as to whether to merge the two boxes into a single sub-Fodorian module: matching and recognition might therefore be regarded as qualifying for one module or for two.

Fig. 6.12: Search and matching/recognition modular framework



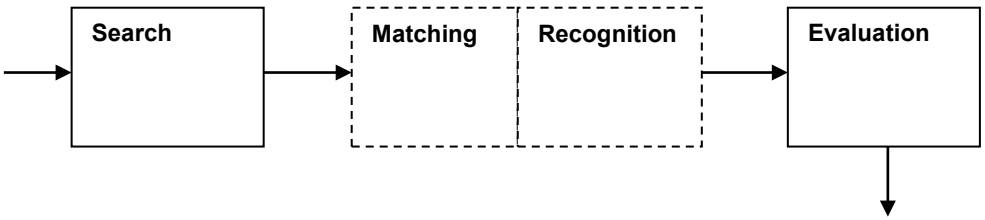
This position is reflected in the diagram above.

The evaluation box accepts input (novel ideas) from recognition. It assesses them by casting them into suppositions, which are combined with the current world model (or part thereof) and then run in the internal possibility space. If the supposition is viable, the idea has potential value and can be considered for expression and onward transmission to other individuals and/or the wider culture.

The supposition and simulation operations might be expected to draw upon internal processing and data, but they also have to have non-Fodorian access to current belief-states, and there is a need to run a possible world simulation. Because of the calls on external resources, the module will only part-qualify as Fodorian, in a similar way to the search module.

Overall, provision is required for a structure of perhaps three or four sub-Fodorian modules at the core of the process.

Fig. 6.13: Core sub-Fodorian modules



Thus the massive modularity structure assumed and taken as a framework from the start of the modelling exercise can be retained, as provided for in the objectives for modelling.

6.6 Counterarguments

The completed model is therefore taken to qualify as a sub-Fodorian modular process, but many other questions as to its validity remain unanswered, most of which will be addressed in chapter seven.

The counterarguments to a Darwinian analogy had to be taken into consideration before commencing the construction of a model, and so were covered in chapter five. The analogy survived this inquiry in a weak form, which left it still in adequate shape for current purposes. The counterarguments could all be accepted without compromising the analogy to the extent to which it has been used in modelling. The model is based around directed search, with variation not decoupled from selective factors, and only approaches the undirected Darwinian ideal where external serendipitous input becomes involved.

Two further lines of counterargument will be addressed at this point. They are derived from the theories of Carruthers and Boden, whose views are more directly opposed to

the model being put forward here; of the two, Carruthers perhaps approaches more closely, and so will be considered first.

Carruthers' theory has been discussed in chapters one and three above. It is set out in his 2002 article *Human Creativity: Its Cognitive Basis, its Evolution, and its Connections with Childhood Pretence*, and in the later book *The Architecture of the Mind* (2006), in particular chapter 5. It is based around childhood pretence, and Carruthers offers strong backing from an evolutionary psychology angle.

As said, Carruthers (2002) suggests that both adult creativity and childhood pretend play require the construction of 'novel suppositions or *imaginary scenarios*' (2002, p.229) and he argues that the function of childhood pretend play is to practise for adult creative thought.

He has creativity as a two-stage process, and draws on the work of Nichols and Stich on pretence. The first stage is the generation of the supposition (by the 'supposer', which is called 'the script elaborator' by Nichols and Stich (2003)). The second is the possible worlds box, used in the evaluation of the supposition. I have adapted these two stages for the evaluation box or module in the above model. Supposition generation is the first stage in Carruthers' view of the process, whereas here I have it as preceded by the search module, which represents the variation that is the source of the novel ideas.

Carruthers' (2002) cites Harris (2000) as regarding pretence as practice for the kind of mental model-building required to understand testimony from others about events elsewhere in time or space. Carruthers points out that pretence actually has an even greater ontological reach (and exceeds the functionality needed for that purpose), as it deals with things out of the actual world altogether. (This extended reach is, of course, suited to constructing models of the possible worlds in possibility space, as I have suggested for my own model, above).

Thus the supposer and the possible worlds simulator in the proposed model are derived from Carruthers and Nichols and Stich. I have taken them up only into the evaluation module (in the selection phase) because they do not seem to constitute a tool for searching or otherwise generating material for variation in themselves, more a means of using such a variant once it has been identified and made available. In other words, they do not seem to have the capability for finding new combinations (in the terms used here, within the possibility space). As said, the search has somehow to be guided, and the stored data has to be structured, to enable practical generation of new variants, otherwise the vast numbers of possible combinations will make the process too inefficient.

Carruthers' (2006) views as expressed in the most relevant parts of *The Architecture of the Mind* have been outlined in chapter three; of particular note are those on massive modularity generally, and on action-first creativity.

The modules I have proposed for the modelled process, even though sub-Fodorian, seem to fit reasonably well with his description of the massive modularity thesis[56]. Their internal processing is isolated from other systems. The search and evaluation modules access information holdings outside the module, but the query and simulation functions should be straightforward and efficient in terms of their calls on other mental processing and information holdings. Searching benefits from the efficiency gains due to direction and to the database structure. Simulation uses the existing belief set, with minor changes, to entertain the supposition.

Carruthers' notion that pretence practices for imagining suppositions about non-actual states of affairs could equally well be applied to the selection stage in my own model, in which the supposition and simulation operations have key roles. Carruthers gives his basis in pretence plausible backing from evolutionary psychology. This line will not necessarily be so readily transferable, but it is less appropriate here because my aim is

[56] See (2006, p.xii).

to show how creative novelty might be originated, without stipulating that any human mental process must be replicated. The evolutionary psychology arguments apply specifically to human creativity; they are not otherwise relevant to the construction of a creative process, provided it is not restricted to the human version.

The model is not action-first, and does not refer to action schemata. This might be taken as an advantage over Carruthers' system, in that the examples of Kekulé and Fleming do not seem to sit easily with the action-first theory, the molecular structure of benzene and the bactericidal properties of penicillin having little immediate connection with actions.

Carruthers refers to associative processes being more tractable than computational process. The above model uses search and query rather than computation, which may be expected to be undemanding in use of resources provided the information to be searched is appropriately structured for efficient access. His description of semantic primitives perhaps more closely approaches searching for required properties, and cuts across domains in the same way as for the searching described here. As mentioned in chapter three, Carruthers' arrangement of associative links between concepts amounts to a structure, and in effect provides for efficient searching.

Carruthers includes a brief description of the way his account would "scale up" in adult creativity (2006, pp.327-330). In this, he does have a directed search, driven by a query raised internally. He also includes reference to a '*similarity heuristic*', seeking something with some resemblance to whatever it is that needs to be explained (in the way that the child's banana-telephone pretence draws on the similarity in shape).

This recalls the search module in the model process, though in that case the search is directed by the requirement, containing the desirable set of properties. The ability to seek and identify similar cases is at least suggestive of structured data[57].

[57] See also the discussion of Carruthers' semantic primitives, in chapter three.

Carruthers also mentions the substitution of linguistic fillers into speech action schemata (see chapter three), and these require a form of search to locate promising candidates. In his view, the supposition is only understood after a filler has been selected; this to an extent corresponds to recognition in the above model, which is when a match would be acknowledged in awareness.

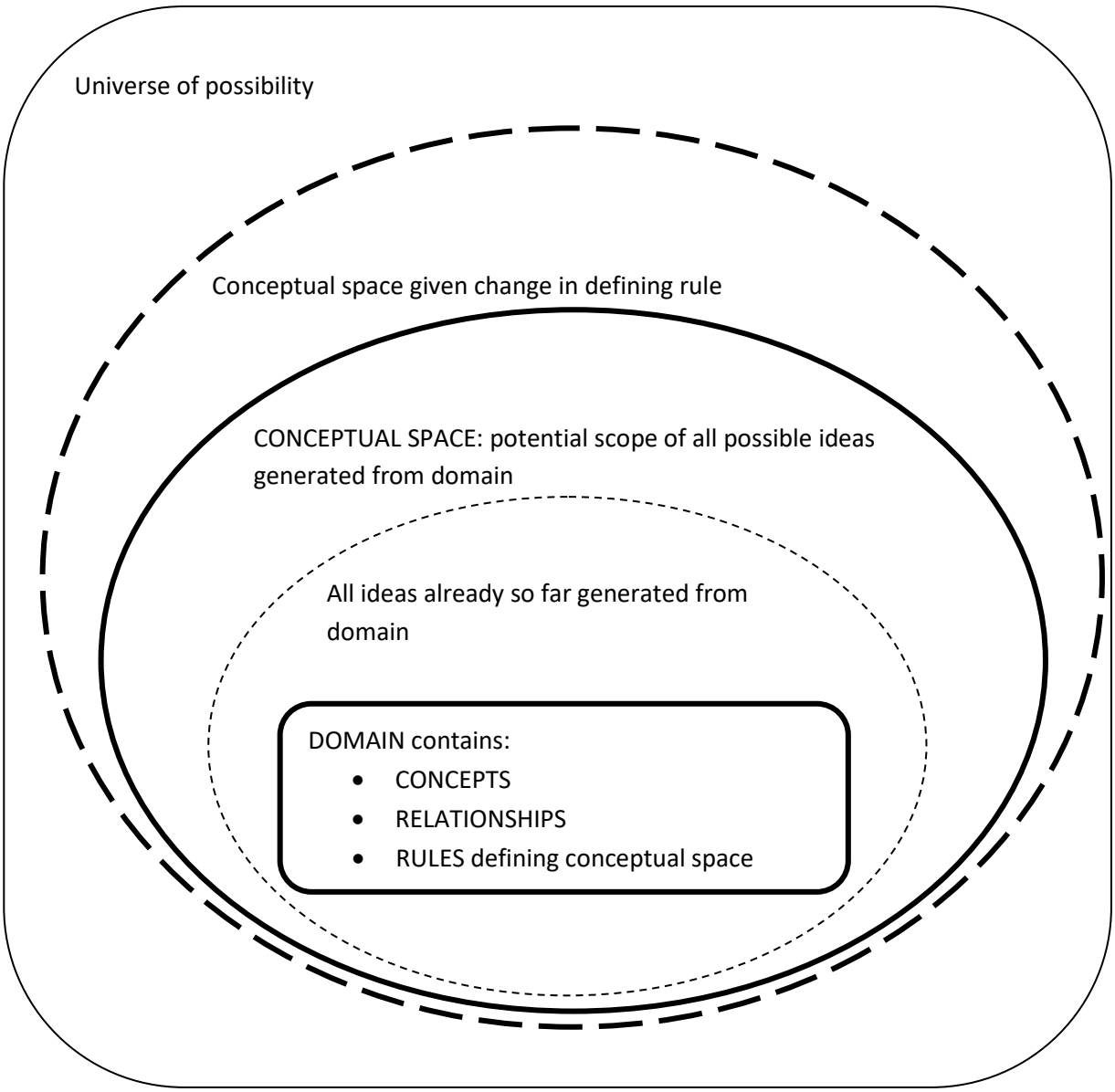
In sum, nothing in Carruthers' scheme seems to run directly counter to the proposed model, and so no modifications or revisions are proposed on this account. There are a number of similarities, even though he does not explicitly include the formulation of a requirement to drive a search. The priming of related concepts is not cast in the form of a query or search, although structure is implied by associative links between concepts.

Boden has a very different approach to creativity, grounded in artificial intelligence and based around conceptual spaces. Her own definition of creativity incorporates the element of surprise as a necessary condition, and she identifies three distinct varieties of creativity corresponding to different types of surprisingness, see chapter two, above. To recap, these are:

1. Recombination of existing elements;
2. Exploring conceptual spaces;
3. Transforming conceptual spaces.

Conceptual spaces are publicly delineated. They are volumes within possibility space with culturally defined boundaries. They set the limits of thought within which creativity in any particular field is traditionally confined, and thus also the range of creative products which is within scope, as long as the boundaries of the conceptual space remain constant and unchanged. I have sketched my view of the relationships between the universe of possibility, conceptual spaces, and domains of thinking, in the following diagram:

Fig. 6.14: Possibility space and conceptual space



The second of these varieties of creativity consists in exploration of an established conceptual space, but the third type involves breaking one or more of the rules defining and constraining the conceptual space, and perhaps replacing them with new ones (which recalls Kant's description of exemplary originality). This has the effect of transforming the conceptual space, thus redefining the area that it occupies in the above diagram.

> We can now distinguish first-time novelty from radical originality. A merely novel idea is one that can be described and/or produced by the same set of generative rules as are other, familiar, ideas. A genuinely original or radically creative idea is one that cannot. It follows that the ascription of creativity always involves tacit or explicit reference to some specific generative system. (1994, p.78).

The constraints imposed by the rules governing the conceptual space (the 'generative system' serve an essential function in structuring fields of thought, although they also tend to confine creative thought and action within their bounds. But the third, more radical, form of creativity transforms and redefines the conceptual space by amending the constraining rule-set in some way. Boden provides examples of dropping or negating such constraints (she later sets up a heuristic called 'consider the negative') including:

> One well-known instance of constraint negation concerns Kekulé's discovery of the benzene ring. … If one considers the negative of an open curve, one gets a closed curve. Moreover, a snake biting its tail is *a closed curve that one had expected to be open*. For that reason, it is surprising, even arresting ("But look! What was that?"). (1994, pp.82-83).

Kekulé's discovery was also taken as an example in the construction of the model suggested here, but in that case the outcome was approached by an entirely different route[58]. The process involved a search of internally stored information, directed by the requirement for string-like items that might form the basis of a molecular structure in line with the known characteristics of benzene. Boden instead starts from the rules defining the boundaries of the appropriate conceptual space, and outlines a process which operates on the basis of amending those rules. This raises a question as to whether there may be more than one type of process capable of generating creative output, and there is no obvious *prima facie* reason why that should not be so. In this instance at least, Boden's approach does not seem to lead to types of creative output which would be beyond the reach of the model suggested above, but other examples given by Boden, such as 'Schoenberg's dropping the home-key constraint to create the

[58] See also the discussion in chapter two.

space of atonal music' (1994, p.82) might not be so readily brought within scope, if at all.

Boden's view of creativity, in terms of conceptual spaces and their associated generative systems, is not the same as that adopted here for modelling. But, there does not seem to be any indication that the existence of Boden's method in any way compromises the validity of the one proposed here. Therefore no modification is suggested, although the likelihood that some creative products might be achievable (or indeed only achievable) by different methods and processes is a significant observation in its own right.

6.7 Conclusion

A high-level boxological model of a candidate creative process has been assembled, in accordance with the guidelines developed in earlier chapters. It sits within a massively modular framework, and is made up of a small number of sub-Fodorian modules. The process follows a weak Darwinian pattern, including directed variation followed by selection. It incorporates additional steps feasible only in human (and perhaps also artificial) settings. The process itself is fully naturalised with respect to its internal operation, and no external inspirational input is required (although serendipitous external influence is explicitly catered for). The model involves working in possibility space, but does not call for the manipulation of generative systems, as is the case for Boden's third type of creativity.

The successful construction of a high-level model cannot not by itself guarantee that it is sound at all levels of detail, nor that it would be capable of implementation as a practical, functioning process. These questions are explored further in chapter seven, which considers to what extent the model might be evaluated and validated.

Chapter Seven: *Findings*

7.1 Introduction

Chapter five was devoted to an assessment of the feasibility of going forward to construct a model of a creative process, based on a modular framework and following a Darwinian scheme, with distinct variation and selection phases. I examined counterarguments from Kronfeldner and determined that there were significant discrepancies between organic evolution and creativity that related to the specific human context, and that would need to be incorporated into the model as additional stages, over and above the basic Darwinian pattern. The feasibility assessment being favourable, and no major obstacle having been identified, I proceeded (in chapter six) to build the model. The construction was concluded successfully, and a model was produced which included the distinctive features identified in earlier chapters. The completed model conforms to a modified Darwinian structure and has a sub-Fodorian modular framework.

In the current chapter, I review and evaluate the degree of success of the model as a whole process. I return to the aims, criteria and benchmarks for the modelling exercise, and consider the extent to which they have been met, taking the process stage by stage. This is followed by an overall assessment of the complete thought experiment, and a consideration of the model as explanation: does it take the understanding of the origin of creative novelty any further forward?

7.1.1 Coverage

As said, the findings from the modelling exercise are assessed in this chapter. It comprises an appraisal of the completed creative process model, together with the techniques employed in its construction, and an examination of the extent to which a

naturalistic explanation of creativity has been supported. These are brought together in an overall rating of the success of the modelling exercise as a whole.

The chapter is divided into sections as follows:

> The proposed creative process and the criteria for success, including an examination of the proposed outline for the model.
>
> An overall assessment, including the methodology and the modular framework.
>
> The process considered as explanation; naturalism and the *Meno* paradox.
>
> Issues remaining unresolved.

7.2 The proposed creative process and the criteria for success.

The objectives for attempting to model a creative process were set in chapter five, and concern the problem of the origination of creative novelty. The principal aim was to explore the feasibility and practicability of mapping out a route for a naturalistic process that could generate output which might potentially qualify as 'creative'. There was also a secondary goal: to seek a naturalistic explanation of creativity by putting together a plausible process capable of generating creative novelty.

There are several assumptions underlying the exercise; some were adopted *ab initio*, and others have emerged as the study progressed. The following are of particular significance when assessing the outcome of the modelling exercise.

> Creativity (or at least one variety of it) proceeds in a manner analogous to evolution by natural selection, at least in the weaker sense that variation is followed by selection or evaluation.
>
> A modular structure can be taken as a framework to support a creative process.
>
> Ideational material does not arise *ex nihilo*, but has to be composed of combinations of pre-existing source elements, just as evolution by natural

selection does not result in the creation of new energy or matter. Therefore, new ideas can only be formed out of material that is already present[59].

It is as if the cosmos contains a universe of possibility in addition to the actual world (as described in earlier chapters) so that in one way a creative act can be regarded as taking something from possibility space and making it actual (as can evolution). There are no ontological claims attached to this assumption.

Creativity can give rise to novel properties, as opposed to new source material (in common with evolution); new ideas can open up undiscovered dimensions in possibility space.

Benchmarks were set up in chapter five, to provide criteria against which the validity and robustness of the model could be assessed. They can be considered in three ways:

1. As to the overall soundness of the model
2. As to whether the model corresponds to a process which can plausibly be expected to be capable of generating creative output
3. As to whether the model corresponds to any form of human creative process, or any part of such a process.

The three modes of application will be taken into consideration when assessing the individual modules and stages of the proposed process. The criteria can be conveniently applied by working through the process in sequence, stage-by-stage, as follows.

7.2.1 Acquisition of prior knowledge and the formulation of the requirement

These preliminary steps generate the requirement, which forms the input to the core modules of the process. There seems little about them which might be taken as at all

[59] See also the discussion of Aristotle's view, in chapter one.

controversial in this specific context. The acquisition of knowledge is an ordinary, familiar mental function, and is not generally held to require anything over and above drawing upon the normal channels of testimony, cultural records, and direct observation. Artificial systems also commonly require access to source data on which processing is to be carried out.

The understanding and formulation of a requirement are, equally, ordinary mental processes and might be expected to be involved in hunting, for example, as discussed earlier on. A simple example might be the compilation of a shopping list, from prior knowledge of items to hand together with those for which there might be expected to be a future need. In the artificial context, a stock control system could presumably be arranged to produce a listing of items and quantities needing to be bought in.

Thus there seems nothing about these preliminary stages which would need to be anything but naturalistic, and they are therefore in principle reproducible artificially. In human terms, knowledge acquisition and requirement formulation might be expected to be part of routine mental functioning right from the early days of the species.

7.2.2 Searching

The search box (or module) takes as input representations of the requirement, and uses these to drive a search, via internally and externally targeted queries. The resemblance between an external requirement-driven search and hunting for prey has already been noted. Serendipitously significant observations might be made at any place or time, and via any sensory mode. This is very similar to the situation as regards prey, for which the identification criteria (the requirement) will be known, but often not the precise time or place of its appearance. The similarity with hunting suggests that the necessary cognitive systems will have been available from early in prehistory.

Internal searching seems more removed from hunting, although retrieval of specific items of memory is an everyday experience. The potential target area, the individual's total possibility space, is unfeasibly large for random searching. Therefore the proposed model suggests restricting the scope by:

1. Using the requirement to define the set of properties sought
2. Directing internal queries at data structured in such a way as to facilitate rapid and efficient searching according to specified properties.

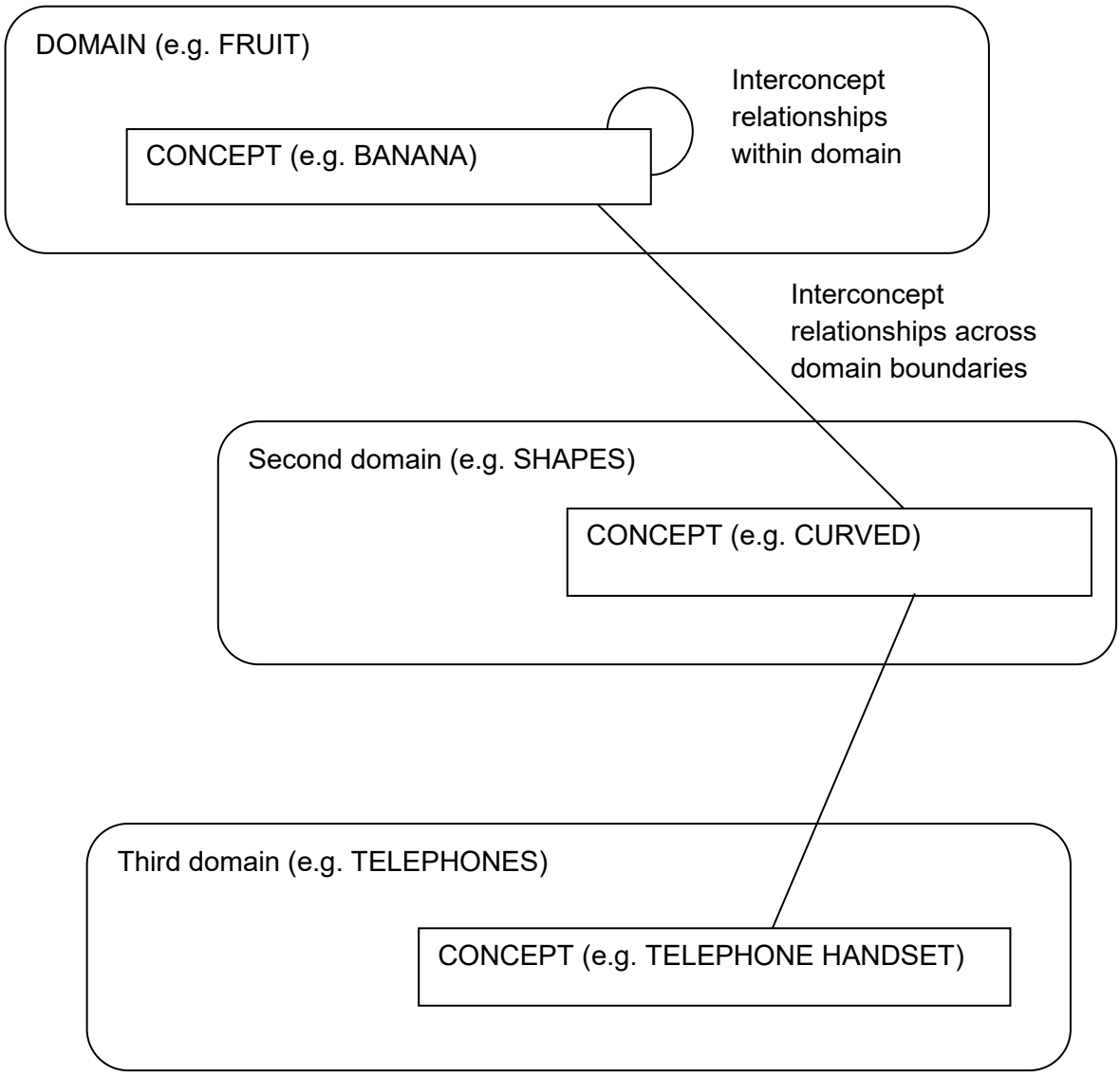
As far as artificial systems are concerned, structured databases and query languages have been available for some years. In the human setting, retrieval according to properties does not seem a demanding stipulation. It is a common experience to recall items of a particular shade of green, for example, and this kind of search is often readily executed. Kekulé's dream of snakes offers an example if regarded as an outcome of memory retrieval of items with properties reflecting possible types of curved configurations. There is perhaps a question as to whether such searching is in practice capable of appropriately efficient retrieval of information according to property set, but final confirmation might need to await experimental investigation beyond the scope of this study. Nevertheless, there is in principle nothing from the above discussions that might be expected to preclude this type of mechanism.

In evolutionary psychology terms, the position on internal searching is less obviously clear than for the external equivalent. One question is whether such searching is carried out on the basis of stored images (in any sensory mode) or of stored conceptual information, which might conceivably be facilitated by the possession of language. If language plays a part, then that would be in line with the very much greater incidence of creative action observed in humans, in comparison with other animals.

7.2.3 Searching within and across domains of thought

If inter-conceptual relationships were to provide links across domains, the resultant lattice might facilitate searching (and analogical reasoning), as in the diagram below, giving a suggested structure for Carruthers' banana/telephone example (see chapter three).

Fig. 7.1: Proposed metastructure of domains



Search queries can access concepts according to specified properties along the inter-domain linkages, which follow properties.

Boden (2004, pp.107-108) refers to semantic nets, and includes a diagram of her own (2004, p.108, Fig 5.4).

7.2.4 Matching and recognition

The matching and recognition module has to compare two sets of data (the requirement and the search outcome) and assess goodness of fit. On the hunting analogy, this would be equivalent to identifying target prey. This suggests that the operation can be carried out using standard, everyday cognitive processing that would have been available at an early stage of human evolution. Similarly, there is nothing here that would not in principle be achievable in artificial systems.

The only aspect that poses problems in terms of ordinary, naturalistic processes is the registering of the recognition in awareness (equivalent to the experience of a eureka moment). But this is part of the much more general mystery of consciousness; it lies outside the boundary of the proposed creative process itself and thus beyond the scope of the current study.

7.2.5 Evaluation

The evaluation module takes as input representations of candidate novel ideas, the output from the matching and recognition module. It casts these as suppositions and carries out mental simulations to assess the potential for added value. Carruthers' views on supposition and pretence were discussed in chapter six. Simulation is also used in artificial systems.

7.3 The model as a process

The model does not require any processing other than might plausibly be expected to be available within the scope of human cognitive resources, even at a relatively early stage in the development of the species. This is not the same as saying that human creativity operates according to the model, only that there is no obvious incompatibility.

No inspirational or otherwise *ex nihilo* input is required at any stage, and there is no break in the chain of causality. Therefore the logical structure appears to be sound,

and there is nothing in the internal flow of information and processing which seems likely to block or hinder the operation of the process. Thus it seems plausibly capable, on the basis of the internal logic, of producing novel output of value. For this to hold, the assumptions must be valid. Also, the model is necessarily high-level and all the lower-level subprocesses will need to be capable of operating as required for the overall process to work (the efficient functioning of the query processes is of particular importance in this respect).

In sum, therefore, the high-level process can be assessed as viable in principle, and so to that extent the modelling exercise has been successful.

The use of the modular framework and boxological presentation proved to be adequate for representing the process according to the extended staging developed in earlier chapters. Three or four core modules have been sufficient to frame the model. The modules are sub-Fodorian but, as discussed earlier, would fit within a massively modular mental architecture.

7.4 The model considered as explanation

As said above, only straightforward, naturalistic processes are called for in the operation of the modelled creative process. The input to the creative process is standard knowledge acquisition, which subsequently informs the assembly of a requirement for a specific set of properties.

No providential intervention is necessary for the operation of the model. Thus insofar as the model represents a viable mechanism for generating novel output of established value, it constitutes a naturalisation of a creative process. It does not qualify as a naturalisation of creativity, because it is not adequate to support a claim that human creativity takes place through the operation of a process which accords with the model. Nor can it be said that the model is the only form of creative process which is possible.

Nevertheless, given that it is sound, it does describe a possible form of creative process.

The question can now be put as to whether, or to what extent, the model constitutes an explanation of creativity (or, to what extent it is capable of supporting such an explanation). Depending on the degree of confidence that can be assigned to any response in the affirmative, it might then be possible to ask whether the existence of the model is sufficient to dispel some of the mystery surrounding creativity.

Darwinian natural selection explains evolution because it removes the need for a designer or for *ex nihilo* creation. Instead, 'design' (new properties) arises through the normal operation of natural laws and processes. The three aspects of variation, selection and preservation are sufficient for evolution to proceed[60].

If the logic of the process making up the model is flawless, it should resemble a sound argument, or mathematical expression, or statement of logic. Then, if good, true, and correct results are taken as input, the output should be equally good, true, and correct. This, then, would constitute an explanation of creativity in the same way as Darwinism explains evolution. It cannot be a complete explanation for the reasons given above: that it is not necessarily either the only process capable of creative output, nor can it be said that human creativity works in the same way, only that the model is not incompatible with what is known about human creativity.

7.4.1 The Meno *paradox*

As described above, the model avoids the trap set by Socrates' *Meno* paradox by focusing on the requirement – the desired set of properties or criteria – rather than on the entity which embodies them. The requirement can be known and sought, even when the nature and location of the solution itself are still yet to be discovered. Indeed, the solution necessarily remains unknown until a novel idea fulfilling the requirement is

[60] See the discussion in chapter five.

identified, otherwise the result is not new, and therefore not an instance of creativity. It is this that explains the element of surprise stipulated by Boden, itself one feature of the mysterious nature of creativity.

7.5 Overall assessment of the outcome of the modelling exercise

Chapter five included a listing of possible levels of success that might be achieved through the modelling exercise. The actual outcome lies somewhere between numbers two and four on the list:

> 2. A plausible creative process is identified, capable of producing appropriate output, but not necessarily reflecting human creativity.
>
> 4. No specific process is identified, but there is some indication that such a process is achievable.

The inquiry has resulted in the identification of a plausible creative process, by constructing an appropriate process model. The identification cannot be definitively confirmed without further investigation at lower levels of aggregation, as discussed. The model does not necessarily correspond to a human creative process, although no points of incompatibility were found.

The position therefore approaches that of the second in the subsequent list of implications (in chapter five):

> 2. To identify a process satisfying some of the criteria for creative output but not necessarily corresponding to human creativity would constitute a naturalisation, but without demonstrating that all creativity is capable of naturalistic explanation. Human creativity might be exclusively inspirational or metaphysical, even if other forms of creativity were possible and were not necessarily so.

If and when positively confirmed, the model would constitute a naturalisation of at least one possible form of creative process. Even then, it would not on its own exclude the possibility that there might exist alternative paths to creative output, not all of which were capable of naturalistic explanation, or indeed of any explanation at all.

7.5.1 Unresolved issues

Any remaining points of uncertainty refer to lower-level processes, or to their efficiency and practicality. This is particularly true of the operation of the search module and the associated structuring of data. It would need to be confirmed that viable searches can be set up to retrieve properties according to the requirement.

7.6 Conclusion

The modelling exercise has been as successful as could reasonably be expected, at the necessarily high level of aggregation that has been attainable with available resource. Definitive verification will have to await breakdown to greater levels of detail (and ideally implementation), though there has been no indication from the study itself that this would not be achievable. The model is internally consistent and requires no break in the chain of causality. Nor does it call for any conscious or volitional input within the internal operation of the process.

In this chapter I have assessed the result of the modelling exercise in terms of its validity as representing a potential creative process, including the method of construction, the internal logic, and the calls on information resources. In the final chapter which follows, I will examine the outcome of the project as a whole, and of the modelling exercise as a thought experiment. I will consider the potential contribution of the study to the understanding of the origin of creative novelty, its wider implications, and the scope for further research.

Chapter Eight: *Conclusions*

8.1 Introduction

In this final chapter, I assess the outcome of the project: the success of the thought experiment; the implications of the findings; and the scope for further study and research. I reconsider the reasons for undertaking an inquiry into the roots of creative novelty. I isolate and scrutinise the original element of the research and I consider the extent to which the results have fulfilled expectations.

My ultimate aim was to address the enduring mystery of creativity, the question as to where it is that new ideas come from. As long as this is unsolved, there remains the problem presented by the genesis of creative novelty, sometimes in a sense apparently out of nothing and nowhere, and sometimes in a manner not easily explainable by naturalistic means. One way to make progress towards deciding this issue is to attempt to prove the opposite: that creativity can conceivably occur purely through processes operating according to the natural laws of the universe. Thus, I set up the research project with the primary objective of investigating the nature of creativity, with particular emphasis on the origination of novelty. I intended to pursue this aim in large part by means of an extended thought experiment, consisting of a process modelling exercise. I wanted to see whether, given the starting assumption that creative novelty is originated naturalistically, it would be possible to specify a process that would plausibly be capable of generating creative output. The positive identification and verification of such a process would go some way towards establishing a defensible account of creativity, showing how new ideas could conceivably be generated by naturalistic means, with an uninterrupted causal chain linking to antecedent conditions.

This had to be cast into the form of a viable research project, and needed a fixed point from which to begin. The foundation was provided by the modularity thesis, taken as a framework, a basis for the construction of a model. Modularity theory offered a way of understanding mental processes to an extent in isolation, and also presented a structural backdrop against which a model of such a process could be set.

8.1.1 The course of the project

A thorough understanding of the scope and characteristics of creativity had first to be established, given the aim of modelling a creative process. Creativity is a complex topic, crossing subject boundaries, and so I commenced with a general survey of literature and sources, which needed to be wide-ranging at first. I included a brief examination of the spread of opinion on creativity, and of the questions and problems associated with the origination of creative novelty. One approach that had already drawn my attention was that adopted by some Darwinian theorists, because there was a clear parallel between evolution, giving rise to novelty in nature, and creativity as a source of cultural innovation. I examined the environmental selective factors which drive evolutionary change, and the view that creative action is a response to challenge (as noted by Kronfeldner). This line eventually contributed to the notion of formulating a requirement, which is a key feature of the modelled creative process. I also observed that Wallas had drawn up a staged view of creative thinking which had gained wide acceptance, and which offered a possible grounding from which to start the development of a process model. The main block of research for this initial survey was carried out in the earliest stages of the project, but I continued to add to my overall understanding of the field throughout, as new material or links came to light.

The initial survey provided a jumping-off point from which to launch a more focused study of the core nature and characteristics of creativity, recorded in the second chapter. I examined and compared the definitions and descriptions of creativity that had been put forward by a variety of theorists, and noted the ubiquitous novelty and

value criteria, their context-sensitive nature, and their insufficiency as complete indicators of creativity. I considered views of the varieties of creativity and novelty and their ontological and psychological dimensions. I investigated the sources of creative ideas in prior states of affairs, leading ultimately to the development of the key assumption that new ideas must draw upon pre-existing mental elements, rather than arise *ex nihilo*. I researched the ontological aspect of creativity and its relation to modal questions of possibility, actuality, and possible worlds theory, given that a creative product corresponded to the actualisation of a possibility. I also explored the *Meno* paradox and its implication of the impossibility of looking for the unknown. This eventually contributed to the development of the search element of the modelled process, in which the target of both internal and external queries was the required properties, rather than their embodiment.

The study of creativity was followed up with a similar inquiry into modularity, in its original Fodorian form and as subsequently developed by others into the massive modularity thesis. I found, as expected, that Fodorian theory would not be able to sustain a model of a creative process, because the relevant components of Fodor's mental architecture, the central systems, were not modularised, and the Fodorian criteria for mental modules were too restrictive. Massive modularity, on the other hand, offered the prospect of an extended modular scheme. I concluded that, if the strict Fodorian criteria could be relaxed in terms of encapsulation and domain-specificity, in ways that would be necessary for modules to support the wide ranging function that is creativity, a modified Fodorian scheme could be proposed that would have the necessary capability. This preliminary work on modularity is recorded in chapter three.

Having completed the initial literature review and investigation of matters central - and also more peripheral - to the subject areas of creativity and modularity, I moved on to consider the feasibility, given the prior understanding of the field by then accumulated, of proceeding to embark upon the construction of a model of a candidate creative

process. I determined the methodology to be adopted for modelling and the criteria for success. I proposed to base the model on a Darwinian pattern, but with modifications derived from the findings set out in earlier chapters. I carried out a thorough examination of the proposed Darwinian analogy, and the adjustments which I expected to have to make in order to allow for the distinctive characteristics of the creative process, as compared with biological evolution by natural selection. The feasibility assessment appears as chapter five.

I then turned to the modelling exercise itself, the core component of the research, which is recorded as chapter six. I used boxological diagramming as the principal modelling technique. The model was founded on Darwinian theory, with significant changes as noted above. Other contributory inputs were the four Wallas stages and the basic novelty and value criteria for creative products.

I had accepted Kronfeldner's conclusions concerning the weaknesses of a strict Darwinian analogy, and the variation element is guided, rather than undirected, except for the random element of serendipitous clues. I cast the process as being driven ultimately by environmental (cultural/artificial as well as natural) challenge. My primary innovation here was to recognise the difference from evolutionary change due to the influence of a creative agent. The creative agent is able to assess environmental pressures and opportunities and to formulate a requirement, a set of desirable properties that would meet a perceived need. This contrasts with the situation in organic evolution, where unguided variation generates novel variants which are pitched straight into confrontation with their environment. In a creative process, the requirement can be used to drive a search. By targeting the search at known or compiled sets of properties, as opposed to (unknown) embodiments of those properties, I found that the *Meno* problem could be sidestepped.

I added matching and recognition boxes to detect candidate novel ideas for output to selection. For the selection phase itself, I adapted the possible worlds box proposed

by Nichols and Stich (or Carruthers' effectively equivalent tagging mechanism) to set up a virtual testing environment against which the potential of candidate creative products to add value could be assessed. The variation and selection phases, when taken together, allowed me to proceed to the assembly of a complete model, retaining the modified Darwinian analogy as a base.

The model at this stage was structured on a raw boxological skeleton, the boxes being, in effect, proto-modules. I proceeded to recast these into a sub-Fodorian modular framework, which could reasonably be expected to fit into in a massive modularity theory context. The result then satisfied the constraining criteria of being both based on a (modified) Darwinian analogy and also being set against a sub-Fodorian modular network, and the thought experiment was now complete.

I sought counterarguments against which to test the model. Boden and Carruthers had views that were (at least apparently) very different from my own - both had theories as to how creative novelty is generated, but neither of them had worked from a Darwinian analogy. On inspection, some elements of the model could be viewed as implied in Carruthers' scheme, in spite of the different basis. I found it more difficult to reconcile Boden's approach from the basis of conceptual spaces, and so this was accepted as a possible alternative mechanism.

I carried out an overall evaluation of the model as a candidate creative process, recorded as chapter seven. I reviewed the aims for the modelling exercise, the benchmark criteria, and the assumptions made, in the light of the results obtained. I verified the model on a stage by stage basis and I considered the extent to which it might be said to have naturalised, or explained, creativity. The outcome is discussed further below.

In this account, I have presented the events in the approximate sequence in which they occurred, however it needs to be said that the research proceeded in a more iterative fashion than represented above, as might be expected. For example, investigation of

the wider background to creativity, recorded here as chapter one, while mostly concentrated at the beginning of the project, continued from time to time throughout, as new points of interest came to be identified.

8.1.2 Outcomes

The body of chapter eight brings together an assessment of achievements throughout the life of the research project. It examines the outcome of the investigation and the modelling exercise in terms of deepening understanding of the nature, scope, and explanation of the origination of creative novelty and the potential application of massive modularity. It considers wider issues prompted by the findings and raised during the course of the study, and asks whether further inferences may be justified. It also outlines the scope for future investigation and for any potential applications.

Here also are set out the original elements of the research. The model is based on the principle of an analogy with evolution by natural selection, in a classical Darwinian sense. But it is not a simple, direct crossover or export from one level of ontology to another. There are crucial departures from the evolutionary template and it is these differences that are central to the functioning and explanatory power of the modelled process.

There are risks, particularly arising from the cross-discipline nature of the topic, and from working assumptions that had to be made along the way. Also, because creativity is closely associated with a number of other concepts, each with their own uncertainties and controversial aspects (and given the limitations of project resources) it was necessary to take strong lines on the meanings to be assumed or adopted for such terms. These various caveats are discussed in full later in this chapter.

Finally, there is an assessment of the overall impact and implications of the results obtained, and a statement of opportunities opened up for further investigation and research.

8.2 Summary of conclusions and implications

The principal outcome of the research project is that a model has been constructed, and put forward as representing a simple process which is conceivably (and plausibly) capable of outputting creative novelty. The model is based upon a modified analogue of Darwinian evolution by natural selection, with a number of significant additions, which are necessary to allow for the distinctive features characteristic of the generation of creative, as opposed to biological, novelty. In common with evolution by natural selection, it incorporates variation, selection and preservation. But it differs from biological evolution in that the variation is *directed*. This is only to be expected in the case of human creativity, because, whereas natural selection immediately confronts external challenge, the embedding of the creative process within a (conscious, critically thinking) human agent allows for the environmental situation to be analysed, leading to the formulation of a requirement, a desirable set of properties. The modelled process is not necessarily the only one capable of originating creative novelty, nor does it necessarily correspond to human creative thought. But it does demonstrate that such a process is conceivable. Also, it does not depend upon inspiration from any supernatural source, nor is any *ex nihilo* origination required, in much the same way that evolution by natural selection needs no designer. But even though there is no call for intervention, there is scope for the advantage conveyed by the influence of a designer.

In addition to modelling a creative process, the outcome of the research project contributes towards an *explanation* of creativity. Explanation is a complex topic, but here it is to be taken in a limited sense, and much as evolution by natural selection might be said to explain the origin of novelty in nature. Natural selection shows how the origination of biological novelty can be possible without any design input, even though living creatures display the combinations of structure and function appropriate to their environment, in much the same way that they might have been expected to had

a designer been involved. Similarly, in the absence of any known creative process, creativity has the appearance of a mystery, because new ideas can seem to come from nowhere, without immediate antecedents, just as if they had appeared *ex nihilo*. The postulated creative process demonstrates a route by which creative novelty might plausibly be originated by naturalistic means, in the absence of any intervention or metaphysical freedom, and preserving the chain of cause and effect.

As regards the Darwinian analogy, it is concluded that the differences between the cultural and biological expressions are as significant as the similarities, if not more so. The modelled process does not call for any external intervention (in that sense it is encapsulated relative to any such input), but does allow for an element of influence from the agent. In particular, the ability to direct a search is in stark contrast to the undirected variation in biological evolution. Here the efficiency gains at the cultural level become evident.

It is the capacity of the agent to compile a desirable set of properties, and then to direct a search based upon that requirement, that bypasses the *Meno* paradox, as discussed in chapters two and five. It may not be known what is sought, or where to look for it, but the required properties can be understood, and can be used to set up a search.

The other crucial difference relates to prior knowledge. In evolution, prior information accrued during evolutionary history is encoded in genetic material, so that evolution is cumulative, but this hard wiring also acts as a constraint, restricting the scope of future change and adaptation. In the process model, prior information is held in a structured network, which allows for associational links between words, concepts and images. This supports the internal directed search, and circumvents the problem set by the vast extent of the possibility space defined by all the potential combinations of ideational elements (which in evolution is constrained by the cumulative effect of previous changes). Searching can be expected to be relatively rapid and efficient, given that it is facilitated by the structure.

The model therefore represents to an extent a reflection of the same process as that of evolution by natural selection, but is a development from it rather than a close analogue, or any simple implementation of an identical mechanism at a cognitive, mental, or cultural level. It is not necessarily the only possible creative process, and its relationship to human creative thought and action remains unknown (although there are no obvious points of incompatibility and there is nothing involved in the modelled process that seems at all likely to exceed human mental capacity).

The straightforward, naturalistic nature of the postulated process prompts the question as to whether it might find an application within the field of artificial creativity; this is discussed further later on.

One final, more general point of difference between evolution and creativity is that the *Meno* problem (see chapters two and five, and also below) does not have any bearing on Darwinian processes, even though novelty can be generated thereby. The reason is that natural selection does not involve a *search*. It operates by natural (blind) variation, and by trial and error. There is no question of anything looking for anything else, and so no pre-life memory need be invoked. The paradox certainly does apply to the generation of creative novelty, and the *Meno*'s reference to searching for the solution can be taken, in its own way, as an acknowledgement of the place of searching in human innovation, and to that extent an endorsement of its inclusion in the modelled process[61].

In sum, the model emphasises the differences from evolution by natural selection, and allows for creation with, or by, design. It retains a basic underlying structure (of variation plus selection and retention) in common with natural selection, and corresponds to a similarly simple and straightforward process. It is not necessarily the only creative process possible and its relationship to human creativity is unknown. On

[61] See the discussion in chapter two.

the other hand, the process does not call for anything of which the mind might be expected to be incapable.

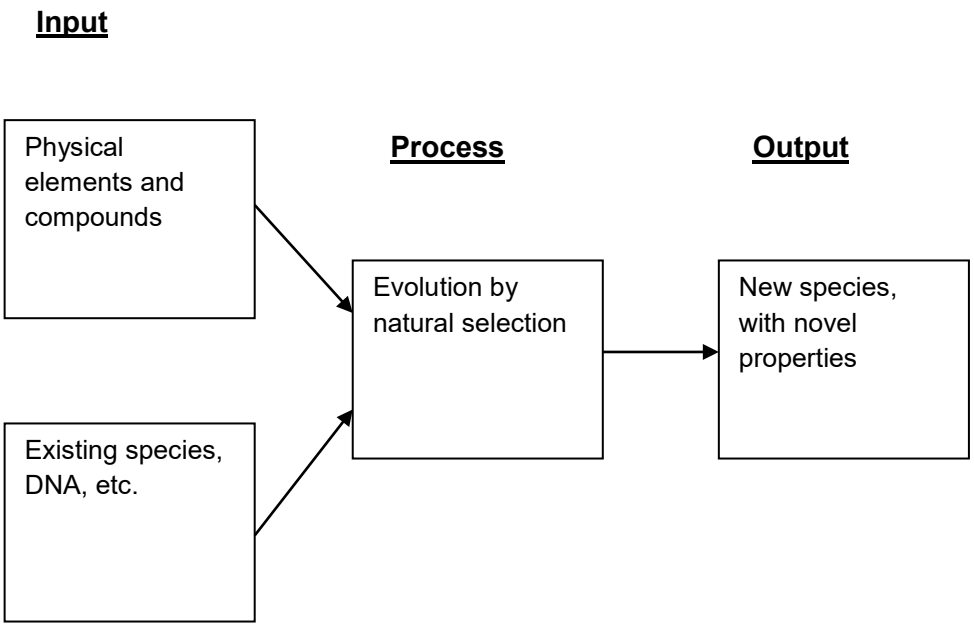
The next section has a direct comparison between evolution by natural selection and the modelled process, to bring out the points of correspondence and deviation.

8.3 The modelled creative process compared against natural selection, by stage

As said, in terms of the Darwinian analogy it is concluded that the differences between the cultural and biological expressions are as significant as the similarities, if not more so. This section examines and closely compares the two processes on a step-by-step basis in order to highlight the points of divergence, and to assess their impact on the functioning of the model.

Evolution by natural selection is a simple, two-step (or three-step, including the retention of selected variations) process, and can be represented boxologically as in chapter two:

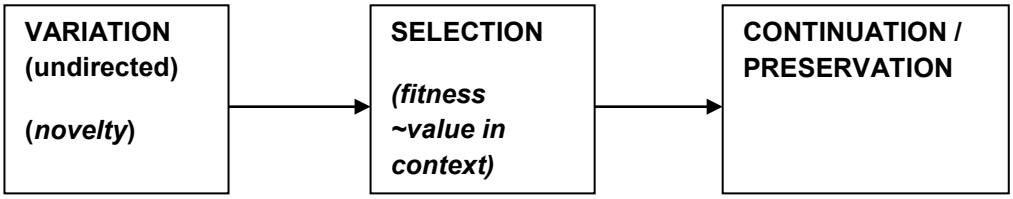
Fig. 8.1: Evolution by natural selection (from fig. 2.6)



There is an unbroken flow in material terms from inputs to outputs. New properties may be generated, but no addition of physical substance or external design input is called for at any point in the process.

The next diagram, taken from chapter six, separates out the variation, selection, and preservation phases:

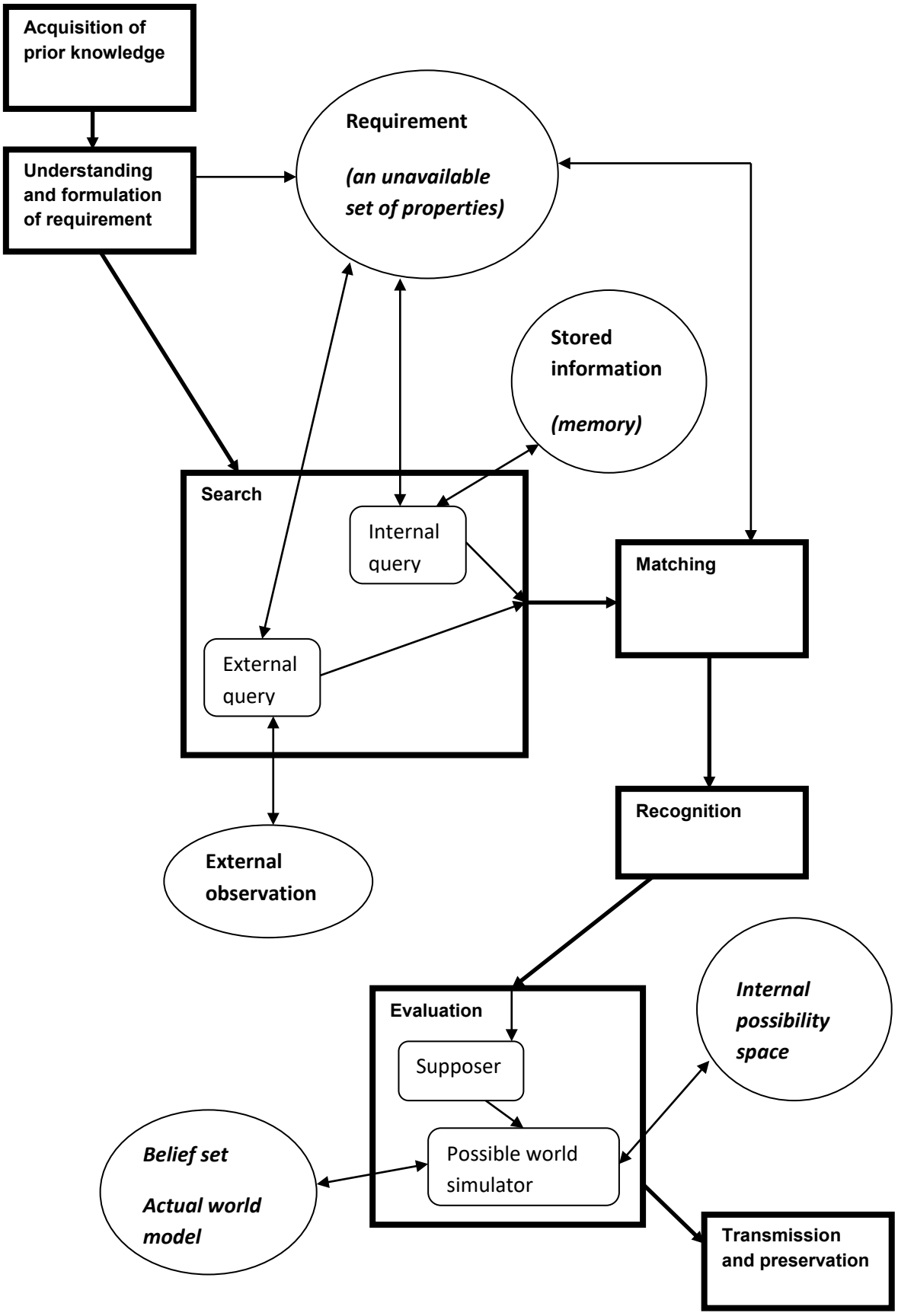
Fig. 8.2: Natural variation and selection (from fig. 6.4)



The three phases consist of: natural variation (through genetic mutation, for example); selection by exposure of the new variants to the environment; and preservation of successful variation (through genetic encoding).

To aid comparison, here is the complete proposed creative process model, again taken from chapter six:

Fig. 8.3: Complete process model (from fig. 6.10)



Here, the inputs and outputs are ideational rather than material. All inputs are sourced from pre-existing mental content. There is provision for design, and for serendipitous observation, but not for intervention or *ex nihilo* generation of creative novelty.

The basic three-step structure is retained; it underlies the additional features of the modelled process and can be picked out from among the various parts of the diagram. The input to the modelled process is split between perceptual data from observation of the external world, and prior information in the form of the internally held structured memory-data which forms the target for internal search (and is also used to formulate the requirement). The variation element is supplied from the search box, and is represented by the outcome of internal and/or external queries. Selection is provided by the matching, recognition, and evaluation sequence. Retention occurs through transmission and preservation or recording (in this case by artificial or social means rather than genetic coding).

There are also features which natural selection does not possess, and here the points of variance from the Darwinian scheme are more meaningful than the similarities. The principal differences relate to: structured data; the ability to formulate a requirement; the facility for directed searching; recognition and matching; and also the opportunity for virtual testing.

Prior information is held in a structured way in biological evolution, encoded in the genetic material for example, but in that form it is inflexible and constrains future variation as well as enabling it. It relates to successful variation in past environmental conditions (see chapter five). In the modelled process, the prior information is also structured (in memory), forming a data repository that can be queried and searched. This might be conjectured to be comparatively mutable, with multi-dimensional webs linking associated information; it might be expected to be much less rigid than the accumulated information hard-coded into genetic material. It is this structure that renders the search efficient (see chapters two and five). As said previously, it seems a

reasonable assumption that there should be such a structure to human information storage given that related material can be readily accessed, for example, a word can be used to retrieve items associated with either its meaning or the written symbols in which it is expressed.

The chief difference, and the defining one to emerge from the research project and the modelling exercise, is the construction of the requirement, which relates to the current environment. The requirement, it will be recalled, is a specification for a set of properties (not currently available in that combination) which will fulfil an observed need or produce a desired effect. There is no direct equivalent in the strict Darwinian scheme. Biological systems are exposed to environmental factors which present challenges to their survival and success. The challenge is a kind of negative analogue of the requirement: a variation carrying a set of properties appropriate to respond to a particular set of circumstances (in other words, fitted) is likely to have corresponding success. But this can only be achieved by trial and error: genetic material has no means of analysing and understanding any environmental pressures with which it might be confronted. The accumulated genetic information limits the possibility space to which the trial and error is applied, while at the same time restricting the range of response.

In the modelled process, the agent's prior information on environmental factors feeds into the assembly of a requirement, an acknowledgement of a set of properties that would be appropriate in a particular situation if only they were available in a suitable form. This permits the response to be targeted, as opposed to straightforward trial and error. The constructed requirement can then be used to set up a search for the specified set of properties.

The other innovation in that part of the model corresponding to the variation element in the strict Darwinian scheme is represented by the search box. The search facility allows for querying the (structured) stored information available in memory, directed by

the requirement. In other words, the search looks for occurrences of the required set of properties, not necessarily restricted to the domain of thinking to which the requirement applies. A watch can also be set up for occurrences of the required set of properties in incoming sense data, allowing for semi-random serendipitous input. Neither external nor internal searching features in evolution by natural selection, where variants arise in the absence of any direction.

In the modelled process, search outcomes are compared against the requirement, so that close matches can be recognised. On the Darwinian scheme, this corresponds to the success of a relatively fit variant, whose properties are a good match with the demands of the environmental conditions[62].

Moving on to the part of the model corresponding to the Darwinian selection phase, the main point of difference is the facility in the modelled process for the conditional construction of possible worlds within the agent's personal possibility space. This allows for virtual testing of the matched solution against a mental model of the actual world, adjusted to allow for the introduction of the novel idea. Thus is enabled the initial trialling of the proposed change, so that a preliminary assessment can be made of the likely viability of the new idea, without the risks of immediate, untested, exposure to the (in this case, generally cultural) environment against which new biological variants are unavoidably pitched, whether they are potentially fitted or not. This adds an element of efficiency to the modelled process. (The selection phase of the model is of course based on previous work by Carruthers and Nichols and Stich[63]).

In sum, therefore, the modelled process preserves the basic three-step sequence of the Darwinian process, but with additional features promoting flexibility and efficiency. They are: the structured prior information, the requirement, the search box, matching, recognition, and virtual testing.

---

[62] Attention was previously drawn to the importance of recognition by Nickles, see chapter two.
[63] See chapters three, five, and six.

The next section moves on to consider the overall position of the natural selection analogy, given the above changes.

8.4 The Darwinian analogy

The discrepancies listed above prompt the question as to whether the analogy with natural selection still holds for the modelled process to any great extent (or indeed at all). The Darwinian scheme was the obvious place to look for clues as to how a creative process might be put together; evolution by natural selection generates novelty in biological nature, and it is unusual among natural processes in apparently delivering something out of nothing. This is also true of creativity, which delivers new ideas, again apparently out of nothing.

As Campbell pointed out (see chapter two), Darwinian evolution proceeds by 'blind variation and selective retention', and so Darwinian analogy theorists have sought to show that creativity proceeds in a similar way. Kronfeldner has found such attempts wanting, as discussed in chapter five. If 'blind' is interpreted as 'random', then the *Universal Library* situation applies, and the sheer size of the resultant possibility space rules this out. It is therefore unlikely that any strict Darwinian analogy will be satisfactory. Instead, it seems apposite that, if a creative process is to bear any relation to Darwinian evolution, it can only be in a modified form.

Natural selection is driven by environmental factors, which determine whether a variant is fit, and thus likely to be successful (that is, its propensity for success[64]). In creativity, the 'environment' represents the (cultural and material) context against which the novel idea is being viewed, and the added value attached to the new idea reflects a perception of the appropriateness of that idea in relation to the environmental context. The added value therefore stands equivalent to fitness, the propensity for success, in natural selection. It will be recalled that Kronfeldner includes a reference to

[64] See chapter five.

environmental factors in her definition of creativity (see chapter two), bringing out this facet of Darwinian equivalence.

In the model, the postulated requirement is a reflection of the challenge presented by some aspect of the environment, as understood by the agent, although there is no direct parallel to the requirement itself in Darwinian evolution because genetic material is not empowered to assess the environmental challenge, only to present variations on the accumulated data from earlier evolutionary change. In evolution by natural selection, influences bearing on variation are (at least partially) causally separate from the environmental challenge, see the discussion in chapter five. In the modelled creative process, the agent is enabled to engage with the context, and to formulate a requirement on the basis of perceived environmental challenge. This is then used to direct the search (targeting the required set of properties), so that the selective (environmental) factors are explicitly brought to bear on the variation aspect, and the causal independence is broken. This is a deviation from strict Darwinism, but it is a strength, not a weakness, of the creative process relative to the Darwinian template, as the resources devoted to generating variants can be deployed more efficiently. There remains a random element contributing to the variation, in terms of the opportunity for input from serendipitous observation.

The other deviation at the variance end of the process is the availability of prior information structured by associative links, as a target for internal search. As said above, this replaces the hard-wired genetic information in biological evolution, which holds the cumulative coded data from previous evolutionary change. That is (at least relatively) fixed, whereas the associative links forming the structure of stored memory data might be presumed to be more open to amendment as experience might prompt. This, then, is a further potential efficiency gain relative to the pure Darwinian model.

These differences in the variance phase resolve some of the deficiencies noted by Kronfeldner in the Darwinian analogy, because although the basic variation and selection sequence is maintained, variation is not blind.

In the selection phase, there are further efficiency gains, as mentioned in the previous section. The facility for mentally modelling possible world scenarios incorporating the novel change allows virtual trials prior to implementation in an actual world situation. This removes some of the risks which immediate live implementation would entail, and which are faced by novel variants in the biological sphere.

Overall, the modelled process permits the resolution of another problem noted by Kronfeldner: strict Darwinism operates by natural law, in the absence of input from a designer. Humans are designers, and creative process needs therefore to allow for design input from the agent. In the modelled process, this is achieved by: the requirement, directed search, structured prior information, and virtual testing; these are specific to, or under the control of, the agent. Accordingly, the innovation is personalised, very much a product of the agent as an individual, rather than of the process alone.

In sum, there is an element of correspondence with evolution by natural selection, but the analogy only holds so far and no further. There are significant differences. In particular, variation is directed, or at least it involves directed search. The key facility for formulating a requirement and directing a search is in stark contrast to the (blind) variation in biological evolution; here the efficiency gains at cultural level become evident. Also, initial evaluation can be carried out by simulation. In particular, the modelled creative process retains the influence of design and designers – a key point of divergence from organic evolution.

The next section takes up these and other issues (such as the *Meno* paradox) having a bearing on the modelled process, and asks to what extent the process itself might be considered new.

8.5 Is the modelled creative process original?

Here, then, a relatively simple process has been postulated for the generation of creative novelty. It has much in common with evolution by natural selection, without being a straightforward analogue. It incorporates the variation, selection, and preservation phases characteristic of the strict Darwinian model. It does not call for external intervention in the form of inspiration sourced from without, nor for any *ex nihilo* appearance of ideational material, just as evolution by natural selection does not require a designer. But it differs from evolution by natural selection, in that it is *directed*. Here is the main focus of influence emanating from the designer, promoting the advantage conferred by design input. This is the key point of divergence from theories of creativity that depend on a strict Darwinian analogy. The modelled process emphasises the differences from evolution by natural selection, and, if taken as a template for human creativity, it allows for humans to create with, or by, design.

In the variation phase, the design input is channelled through the direction of the search, which is in turn driven by the requirement. The requirement is a desirable set of properties, the assembly of which is informed by the prior knowledge of the (natural and/or cultural) environment. This is the aspect of the modelled process that circumvents the *Meno* problem; it can do this because the environmental challenge and corresponding requirement can be identified and understood in advance, even when that specific set of properties is not as yet known to be available. The occurrence of the target set of properties can then be recognised, even when their embodiment has not been encountered before.

The model is still firmly based on a Darwinian analogy, the use of which is, as discussed earlier, by no means original, having been previously put forward by a number of theorists. But in the modelled process, the Darwinian analogy is adopted in a much modified form, the deviations from the Darwinian base being taken as integral to the modelled process, rather than as signs of shortcomings in the analogy. The

differences can be regarded as indicating the strength of the Darwinian scheme, in being capable of conversion and upgrade to handle creative process as well. The departures from the Darwinian standard mean that, taken as a whole, no equivalent model based on a Darwinian analogue has been put forward before, as far as the author is aware. However, elements of the modelled process certainly have appeared previously, examples of which are given below.

The closest approach to the process as set out above appears in Carruthers' *Architecture of the Mind* (2006). This has, or implies, some similar features, although Carruthers refers to, and draws on, pretence, rather than setting out from any Darwinian analogy. Given the different starting points, the degree of convergence between aspects of Carruthers' system and the variation phase of the model is reassuring.

Carruthers makes reference to a sentence structure and fillers (see chapter three), appropriate fillers being found as the result of a search. The search presumably in turn implies matching, to determine which of the retrieved items is to be used. Identifying a suitable filler also implies its recognition; a selected filler must have satisfied a set of criteria, whether or not the individual indulging in the episode of pretence is consciously aware of this. The set of criteria implicitly corresponds to the requirement in the modelled process[65] (although Carruthers does not emphasise this aspect).

Carruthers' use of 'semantic primitives' (as described in chapter three) as shared between concepts, amounts to a structured information store, with associative linking, so that one concept can 'prime' others related to it by links of similarity, much as a query might be directed along associative linkages in a search of structured information.

[65] For example, Carruthers mentions that, in searching for possible fillers in an episode of pretence, a child '…might attempt to recall some admirable things, activities, or types of people.' (2006, p307).

The selection phase of the modelled process is derived from the use of supposition and the possible worlds box by Carruthers and Nichols and Stich. The modelled process explicitly involves supposition, which Carruthers uses (in effect) to set up the possible world. He actually refers instead to the 'tagging' of downstream beliefs (see chapter three), and explains that it could be regarded in one way as the equivalent of constructing a possible world.

Boden's treatment of creativity views it as having combinational, exploratory, and transformational varieties, depending on the element of surprisingness (see chapter two). These have different mechanisms, as described earlier. In the most radical, transformational, form, the new idea is arrived at by an apparently thoroughly dissimilar route: the transformation of the appropriate conceptual space by modifying its defining criteria or rules. The identification of a suitable candidate rule modification will require recognition, and presumably also a search, but as Boden describes it the search is carried out on different principles from the directed querying of structured information envisaged in the model. The search and trials centre on the generative system defining the conceptual space, rather than on stored prior information. Boden does, however, refer to 'semantic nets' with a linked structure (see chapter seven). Boden's scheme has not been studied in depth here, and must therefore be accepted as a separate, alternative mechanism (although see also the comment on her approach to Kekulé, in chapter two).

In sum, the model presents a high-level creative process based on a modified Darwinian analogy and set against a massively modular framework (as below). The new developments from the basic Darwinian scheme are the features that enable creative novelty to be generated. In particular, the structured prior information, the construction of the requirement, the facility for directed (as opposed to blind) query and search, and matching and recognition, are all additions to the Darwinian base. As explained above, where they can be discerned (whether explicitly or implicitly) in the

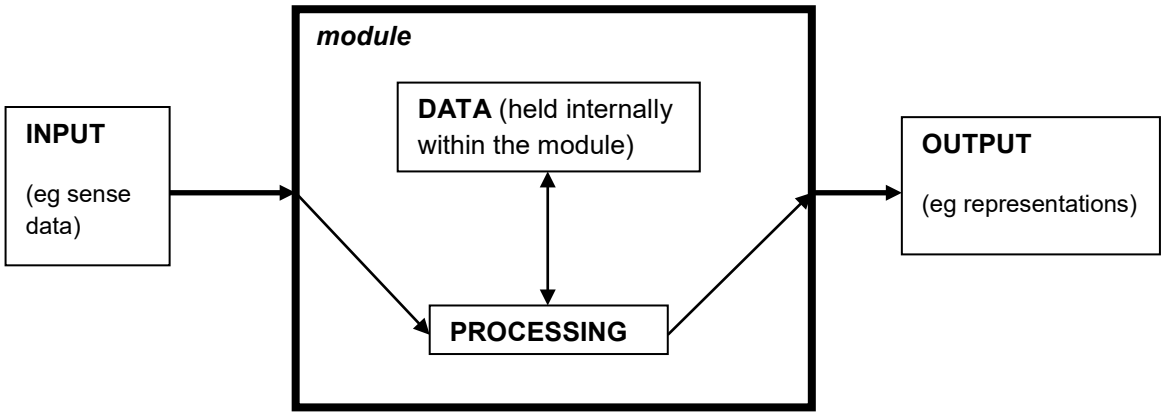
writings of other authors, their use in this specific context is generally new. The principles of searching and of the structured information base are foreshadowed in the work of others, as above, and the model could be regarded as to an extent a fusion of certain elements from Carruthers' *Architecture* with some derived from Darwinian analogy approaches (via Kronfeldner), but other features are unique to this model.

8.6 Modularity

It is the modular structure that has provided the architectural background enabling the formulation of the model. This framework is necessary to allow the process, the series of steps leading up to the generation of creative output, to take place. Fodor's holistic central systems would not support such a process (creativity would have to be mediated by some other means). Therefore the Fodorian modular system could not be adopted unaltered for current purposes. Instead, a type of massively modular arrangement, allowing for (at least partly) modularised central systems has been taken up. Accordingly, the individual modules in the model are sub-Fodorian, some of Fodor's criteria having been relaxed to allow for the necessary functionality.

To recap, the basic Fodorian module was sketched in chapter six as follows:

Fig. 8.4: Classic Fodorian input module (encapsulated, domain-specific), from fig. 6.1



As explained in chapters three and six, the modules cannot in general be made to conform to the key Fodorian criteria of encapsulation and domain specificity. This partly arises from the nature of creativity, which is a much more wide-ranging task than the relatively straightforward, peripheral, input and output functions handled by Fodor's original modular scheme. Creative ideas can require the application of principles taken from one sphere of thinking to another, not normally closely associated, domain (Kekulé's daydream being a case in point). Therefore, a creative process cannot be constrained to be domain-specific in the classical, Fodorian, manner. It might perhaps be considered to have relative domain-specificity by virtue of the fact that the modelled creative process is dedicated to one particular type of mental activity.

Similar considerations apply to encapsulation. The creative process modules simply cannot be encapsulated in the strict Fodorian sense, because of the need to access external data at certain points in the process. This contrasts with the classical Fodorian input and output modules, where no external data needs to be called upon (and indeed Fodor makes a good case as to why the intramodular processing must be kept free from external influences). The creative process modules have to have

access to external data: the constructed requirement; sense data (for serendipitous observation); and the internal actual world model (for building possible worlds). But this is not the same as the direct influence which Fodor barred from his own modules. Accessed data will not directly influence the processing as such. It will affect the sourced input, but then the choice of input data is not within the control of a Fodorian module, either; only the processing is confined entirely within the module boundary. Therefore, the modules in the model might be considered to have semi-encapsulation, partly along Fodorian lines.

Thus, the modular structure is integral to the modelled creative process, which could not be set up in its absence. The postulated modules are necessarily sub-Fodorian, although on examination, they may be taken as not departing very far from the Fodorian ideal.

This completes the discussion of the components, functioning, and degree of originality of the modelled process. The following sections turn to the overall outcome and achievements of the project, together with the associated risks and the assumptions made, and subsequently proceed to look at implications and at potential forward directions.

8.7 Results and findings

The results of the literature review are set out in the earlier part of the thesis, mainly in chapters one to three. The review concentrated upon areas of specific relevance to the study, including: the nature of creativity; modularity and massive modularity; cosmology, possibility, and actuality; Darwinism; Darwinian analogy theories; pretence; and evolutionary psychology. The review enabled a thorough understanding of the issues surrounding the origination of creative novelty, the results of which were fed into setting up the modelling exercise.

The chief defining qualities of creativity were found to be the generation of cultural novelty, in conjunction with added value, both of which referred to the product, or output, of an instance of creativity. The process by which any such product might be brought forth was problematic, largely because new properties, not present prior to the creative act, could be brought into being in the absence of any obvious link to antecedent conditions. In addition, the generation of a creative product as a fresh combination of pre-existing ideas posed problems because of the large number of possible such combinations for even a relatively small number of source components. The closest naturalistic parallel to creativity was provided by Darwinian evolution by natural selection, which generates new properties in nature.

Theories of modularity afforded a basis for understanding creativity as a process, that is for constructing a series of steps that might be expected to lead to the generation of creative output. Massive modularity offered a theory of mental architecture potentially capable of supporting a creative process, and had indeed been put forward as such by Carruthers. Evolutionary psychology was found to be important in establishing a case for modularity, although the identification of a candidate creative process on modular lines would not in itself imply that human creativity was mediated through that same process.

The review of literature had provided a degree of confirmation for basing the model on a Darwinian analogy, for which no sufficiently promising alternative had been identified. The Darwinian analogy was by no means a straightforward option, because it had become clear that evolution by natural selection could not be mapped directly onto a process for generating creative novelty. Accordingly, a detailed examination was carried out of the case for the analogy, and of the counterarguments which might be taken to oppose its adoption. It was concluded that the analogy was only applicable in its most basic form, but that this would be sufficient to proceed to the modelling exercise.

The completed model accordingly corresponded to a high-level process based on a modified Darwinian analogy, and conformed to the intended sub-Fodorian modular framework.

The construction of a high-level model for a creative process, to the extent that is successful and reducible to deeper levels of detail, demonstrates how creative novelty might possibly be generated. The process is therefore a plausible candidate for implementation in an artificial system, at least. It is also a candidate to represent human creative process, creative novelty arising exclusively (or almost exclusively) through human agency. However, there is no indication that the model depicts the only possible creative process, and so there is equally no basis for presuming that human creativity is mediated by any equivalent means. There are, however, no obvious grounds for thinking that the modelled process could not possibly be carried out within the mind, given human mental capacity and evolutionary development.

The principal end result of the research project is the establishment of a plausible creative process, which is not ineligible for consideration as a candidate process for human creativity. Insofar as the model is successful, and reproducible at lower levels of detail, it also offers an explanation as to how the origination of creative novelty might be possible by naturalistic means, without recourse to *ex nihilo* generation or external, inspirational intervention. This corresponds to a limited sense of explanation, and is equivalent to the way in which Darwinian evolution by natural selection presents a process explaining how novelty can arise in biological systems by naturalistic means. It cannot by itself provide an explanation of human creativity, for the reasons given above, although it does indicate one route through which human creative thought might possibly arise.

8.8 Assumptions and risks

A number of assumptions have had to be made during the project, and the study outcome has not been reached in the absence of risk. The main assumptions made along the way, and the principal risks, are set out below.

Assumptions

The following is a list of the main specific assumptions made at various points during the study.

1. It was declared as a basic assumption (in the introduction) that creative processes could be investigated in the reasonable expectation that explanation *is* achievable.

2. It was assumed that creative products are the effect of antecedent causes that are (in principle) knowable, and that a creative process does not require intervention, these being conditions for explanation to be feasible (from chapter one).

3. The universe of possibility was assumed as the set of all points in all possibility space (from chapter two). This was proposed as a construct for the sake of argument.

4. It was assumed (in chapter two) that creativity does not involve any break in the causal chain, even though new properties may arise that were not present in any prior input state. This was related to a second assumption, namely that:

5. creative output is sourced from elements of pre-existing mental content, whether internal stored memory or representations of current sensory impressions – in other words that ideational material does not appear *ex*

nihilo. Creativity *can* give rise to novel properties, as opposed to new source material (as also does evolution).

6. It was assumed in chapter five that evolution by natural selection can be taken as a sufficiently good analogy for creative process (for at least one form of creative process, and in the weaker sense that variation is followed by selection), for the specific purpose of constructing a model.

7. Prior information is held in a structured way in biological evolution, encoded in the genetic material for example, and in chapter seven it was assumed that the prior information drawn upon in the modelled process is also structured (in memory), forming a data repository that can be queried and searched.

8. It was assumed from the start that a modular structure can be taken as a framework to support a creative process (see chapter three).

These assumptions were discussed at the relevant points in the text. Numbers five to eight in the above list are, of course, particularly fundamental to the construction of the model, and are to a certain extent justifiable by results, in so far as the modelled process is found to be plausibly capable of generating creative novelty.

Risks

The chief element of risk derives from the cross-disciplinary nature of creativity. Along with cognitive science, it straddles the boundaries of philosophy, psychology, and neuroscience. In addition, creativity borders artificial intelligence and evolutionary psychology. And, of course, the results of creative thinking are observed in science, the arts, technology, and just about all other fields of human activity. This immediately introduces elements of risk because principles established in one discipline might not apply, or not apply in the same way, in another. Kronfeldner (2007, p.iii) wisely and correctly makes a similar acknowledgement of cross-disciplinary risk.

One prominent area of risk, as explicitly acknowledged right up front in the introduction, is down to the fact that creativity touches on a number of problematic topics and concepts, each with their own elements of uncertainty. It has not been possible, due to resource limitations, to carry out full analyses of many such areas. Therefore, working definitions have been adopted for the purposes of the current study. The main areas concerned are listed in summary form below. 'Creativity' itself is included for completeness.

> ***Creativity***. Creativity is chiefly recognised by publicly observable output, which must display properties of novelty and added value relative to antecedent conditions. The novelty (for purposes of understanding creative process) is of the psychological variety: it must have aspects that are original, and not copied[66]. The added value is always relative to the context; it is a measure of the appropriateness of the response to environmental challenge.
>
> ***Emergence***. The appearance of a novel property in a whole, which is absent from any of the separate parts.
>
> ***Explanation***. For current purposes, this has been taken in the restricted sense in which Darwinian evolution by natural selection might be taken to *explain* the generation of novelty in biological systems.
>
> ***Model***. A representation (of a process) in symbolic form.
>
> ***Process***. A series of steps, stages, or phases leading up to a specified outcome, product, or result.
>
> ***Requirement***. A set of qualities adequate to fulfil a particular need or set of criteria.

[66] See chapters two and five.

The elements of risk cannot be totally eliminated, of course, but it is impossible to set out to investigate creativity without accepting them. Up to a point, the result helps to confirm the validity of the approach and to justify the acceptance of the risks; given that the modelled process is capable of producing creative novelty, it is to some extent self-validating and contributes towards the justification of the assumptions.

8.9 Implications

The remainder of this chapter returns to some of the wider issues raised in the general survey of the field (in chapter one) and looks at them again, this time in the light of the experience gained in carrying out the study.

8.9.1 The mystery of creativity

This section explores the implications, assuming for the sake of argument that the constructed model accurately represents the only possible creative process, and that all creativity (including human creativity) conforms to the model. In that case, creativity is an entirely naturalistic process, and in principle explainable (or at least, explainable insofar as natural selection explains evolution and the origination of novelty in biological systems).

What then is left of the mystery of creativity? As said previously, creativity certainly has every appearance of being a mystery; new ideas can seem to arrive out of nowhere, and to be disconnected from antecedent thinking. For one thing, even if the modelled process were to be taken as a complete description of the way in which creative novelty is generated, whether through human or artificial agency, there remains the ontological aspect of mystery, that is: the problem of origination, as previously discussed. How do new properties arise from combinations of existing things that do not themselves possess those properties? The modelled process shows how appropriate new combinations of ideational material might be assembled and identified, though it does not demonstrate or explain how it is that novel combinations can display

properties that were absent from the original components. The components are preserved (they are merely rearranged) and the new combinations are entailed by the process. But perhaps the new properties are not so entailed, just as the properties of a bicycle are not present in its individual component parts, and are not somehow manufactured during the assembly of the components into the whole. The process may be naturalistic and explainable but the emergence of novel properties is less readily understood.

Some of the problems centre on the relationship between possible and actual. First, there is the presumption that possibilities exist at all (assuming that actualism is not a true description of the world). This links to the aforementioned question whereby the universe of possibility, consisting of all possible states of affairs, is in some ways the equivalent of nothing at all, and that the actual world seems only to have being on condition that just some possible things, rather than others, are actual. And might this also reflect the mystery that counterfactuals and conditionals are conceivable, in addition to actual states of affairs, which in turn allows for the possibility of creative novelty?

The modelled process preserves the individual, personalised, character of creativity. This derives from the differing prior experience of creative individuals, and hence the uniqueness of their personal possibility spaces. There is also serendipitous observation, which adds an unpredictable, and at least partly random, element, even if it is taken that the external search for the required set of properties is directed by a synthesised requirement.

8.9.2 Designers, virtual testing, and the environmental drivers of creativity

Natural selection takes place in the absence of design. As noted earlier (in chapter one) the driving forces for change and development are the tests presented by the environment. The products of natural selection go to make up the diversity of life. But, since the evolutionary selection factors are also the environmental challenges with

which all individuals and species have to contend, the process that gives rise to them can be unpleasant, as Darwin himself observed. This points to one of the differences, or disanalogies, between creativity and evolution: the human capacity for selection to be trialled mentally, or in an artificial exploratory or experimental setting such as a laboratory, or an artist's sketch pad. These are ways of working in possibility space which are unavailable to purely biological systems. Even if the creative generation of novelty is not itself accessible to the will, as Kant noted, the trialling and introduction – or withholding – of novel ideas can be performed consciously, in awareness of the circumstances, and ideally with a view to the likely consequences (the selection can be directed, as well as the variation). Human creativity and the culture at least have the opportunity to bring in the ethical dimension that Darwin and Mill became uncomfortably aware was lacking in the biological sphere. In addition, the ability to carry out trials in possibility space, rather than the actual situation, adds a layer of efficiency that biological evolution is not equipped to exploit.

Hence the differences between creativity and evolution are of particular significance and are indicative of the extended scope that the development of creativity and culture affords. Natural selection is *driven by* environmental challenge (selective) factors, whereas creativity *responds to* a requirement (ultimately derived from an appreciation of environmental challenges, in this case, cultural as well as natural). Thus the requirement, which drives the search and provides the selection criteria, represents the agent's assessment of the set of properties needed to confront a perceived challenge. This is one of those points of divergence at which the strict Darwinian analogy falls down. The discrepancy has a positive consequence in that, in creativity, selection can be carried out to an extent in the imagination, in mental simulation in which there is not the same element of risk as in natural selection, which is always live and dangerous. In human creativity, mental trial by simulation can be extended, for example to virtual reality situations in cyberspace, with potential for increase in range and efficiency – this greatly enhances the advantage of creativity relative to evolution.

It is beyond the remit of this project to comment on such matters any further, any link with volition being out of scope, but it is an interesting observation that the distinction between the natural and cultural contexts in this case mirrors the types of difference that this study has identified between creativity and evolution. Within the creative process itself, as modelled, the origination of novelty is encapsulated with respect to the wider issues of consciousness and volition, these being unnecessary for the internal operation of the process.

Thus, natural selection is driven by environmental selective factors (which stand equivalent to the requirement in the constructed model). Organic evolutionary processes have no ethical dimension, and so the welfare of individuals is not a factor, as noted by Darwin and Mill. There is a sense in which, in lamenting the suffering observable in nature, Darwin mixes up the two ontological levels or categories, organic and cultural. The facility for initial virtual trials in the selection phase of the model gives the potential to address some of the problems that Darwin himself observed. The creative process incorporates input from a designer, which is how this is brought about, and design only becomes relevant with consciousness and humanity, themselves outputs of natural selection. If human creativity indeed follows a modified Darwinian pattern, then it would seem that evolution by natural selection, that universal generator of novelty, has ultimately produced a faster, more flexible, more efficient and altogether more powerful, process for achieving a similar end. Or put another way, in enabling the origination of creative novelty, evolution by natural selection has transcended itself.

8.9.3 Automation

The modelled creative process does not presuppose any connection with human creativity. Neither does it preclude *ab initio* the implementation of the process on an artificial platform. The model has been constructed as a higher level process, whereas automation would require lower levels of detail to be drawn out, and they in turn would need to prove viable and suitable for automation. However, having acknowledged

these caveats, there seems nothing to prevent further exploration along those lines. The structured information, query, and search facilities, and the generation of the requirement are untested in this specific context in any physical installation, and the success of any such venture would be dependent upon whether these components, when harnessed together, are capable of generating creative novelty, and whether they are able to do so sufficiently effectively. Provided that these potential obstacles could be overcome, and the model were to be implemented onto an appropriate artificial platform, there might be some prospect that the process could prove capable of yielding creative novelty, or at least some simpler form of output which might be shown to meet scaled-down criteria. If this was found to be the case, then substantial progress would have been made on the way to answering Boden's final Lovelace question in the affirmative (though might it be more appropriate to regard the process as generating the creative output, rather than the computer?):

> As Boden [(1994, pp.84-85)] reports it, Lady Lovelace was notoriously dismissive of the claim that computers could have anything to do with creativity. But Boden suggests that we distinguish four questions, which she calls 'Lovelace questions' in honour of Lady Lovelace. […] The fourth question (which is the one that Lady Lovelace had primarily in mind) is whether computers could ever *be* creative, and this question Boden leaves aside on the ground that it is a philosophical and not a scientific question. (Beaney 2005, p.180).

8.10 Further directions and opportunities created by, or raised during the course of, the study.

If the outcome of the project throws any light on the central questions addressed, particularly the origination of novelty, it also points to many further directions of inquiry which naturally lead onwards from it.

It is as if creativity draws from the universe of possibility, only a proportion of which is known or predictable from current knowledge, hence part of the reason for its capacity for surprise. In the universe of possibility (and perhaps even in the *Universal Library*), lie all possible solutions to any problems one can think of, where they exist at all. Thus

creativity, the only means of access to much of this realm, is of immense importance, and anything that can provide clues as to how such access is achieved, or how to facilitate, broaden, or deepen this pathway can only be of benefit.

One area of potential application is in education and training: to the extent that insight into the operation of creativity is improved, there would presumably be advantage to be gained if this could be utilised in training in the development of creativity, although this prospect is clearly very remote.

In the modelled creative process, structured information is the key to efficient searching and access. A possible question here on the evolutionary psychology angle is whether any similar structure applies in the human context, and if so whether the structure goes hand in hand with the development of language and/or concepts.

The model itself might permit of further development and implementation. The process may lend itself to automation on an experimental basis, at a sufficiently reduced scale. It would require a system with access to a structured database, a query and search facility, and the construction of a virtual world with a sufficiently small universe of possibility to be of manageable size for the purposes of the trial.

8.10.1 And finally…

The ultimate goal, above and beyond more modest aims for the current project, is essentially to do for the generation of creative novelty something of the same kind that Darwin accomplished for the origination of novelty in nature. Darwin came up with a process which shows that evolution is necessary as a natural law, given the right conditions[67]. Creative novelty is contingent (upon the existence and will of a suitable creative agent, for example), and not necessary, but the outcome of the study demonstrates that a process is conceivable which is plausibly capable of generating creative novelty. I am also encouraged by the close link with Darwinism. It is

[67] See Nickles (2003, p.56) and the discussion in chapter two.

particularly pleasing that the modelled process shares the basic structure of such an established theory of the origin of novelty in nature. The success of the model supports the view that creative novelty could be generated by means of a naturalistic process. I see the development of such a model as offering a base for further research and debate on the origin of creative novelty; I have already indicated a few starting points.

Bibliography

Aristotle, *The metaphysics*. Translated by McMahon, J.H. Buffalo, New York: Prometheus Books, 1991.

Beaney, M. (2005) *Imagination and creativity*. Milton Keynes: The Open University.

Bennett, J.G. (1998) *Creative thinking*. Santa Fe, New Mexico: Bennett Books. (First published 1964, Coombe Springs Press).

Blackburn, S. (2008) *Oxford dictionary of philosophy*. 2nd edn. revised. Oxford: Oxford University Press. (1st edn. published 1994).

Boden, M.A. (1994) 'What is creativity?', in: Boden, M.A. (ed.) *Dimensions of creativity*. First MIT Press paperback edn. Cambridge, Massachusetts: The MIT Press, (1996), pp.75-117.

Boden, M.A. (2004) *The creative mind: myths and mechanisms* 2nd edn. Abingdon: Routledge. (1st edn. published 1990.)

Boden, M.A. (2014) 'Creativity and artificial intelligence: a contradiction in terms?', in Paul, E.S. and Kaufman S.B. (eds.) *The philosophy of creativity: new essays*. Oxford: Oxford University Press, pp.224-244.

Borges, J.L. (1939) 'The total library', in: Borges, J.L. *The total library: non-fiction 1922-1986*. Translated by Allen, E., Levine, S.J., Weinberger, E. Edited by Weinberger, E. London: Penguin Books Ltd, 2001, pp.214-216. (*The total library: non-fiction 1922-1986* first published 1999).

Borges, J.L. (1942) 'The library of Babel.' Translated by Irby, J.E. in: Borges, J.L. *Labyrinths*. Edited by Yates, D.A. and Irby, J.E. London: Penguin Books Ltd, 1970,

pp.78-86. (First published in: *El jardin de senderos que se bifurcan*. Buenos Aires: Sur).

Boyd, R. (1979) 'Metaphor and theory change: what is "metaphor" a metaphor for?', in: Ortony, A. (ed.) *Metaphor and thought*. Cambridge: Cambridge University Press, pp.356-408.

Burke, J. (1978) *Connections*. London: Macmillan London Limited.

Cain, M.J. (2002) *Fodor: language, mind and philosophy*. Cambridge: Polity Press.

Campbell, D. T. (1960) 'Blind variation and selective retention in creative thought as in other knowledge processes'. *Psychological Review*, 67 (6) pp.380-400.

Carruthers, P. (2002) 'Human creativity: its cognitive basis, its evolution, and its connections with childhood pretence', *British Journal of the Philosophy of Science*, 53, pp.225-249.

Carruthers, P. (2006) *The architecture of the mind*. Oxford: Clarendon Press.

Chomsky, N. (2009) *Cartesian linguistics*: *a chapter in the history of rationalist thought*. 3rd edn. Cambridge: Cambridge University Press. (1st edn. published 1966, New York: Harper & Row.)

Darwin, C.R. (1860) Letter to Asa Gray, 22 May, in Darwin, F. (ed.) (1887) *The life and letters of Charles Darwin.* London: John Murray, vol. II, pp.310-312.

Darwin, C.R. (1996) *The origin of species*. Edited by Beer, G. Oxford: Oxford University Press. (*The origin of species* first published 1859)

Dawkins, R. (1986) *The blind watchmaker*. Harlow: Longman.

Dawkins, R. (1989) *The selfish gene*. Oxford: Oxford University Press. (1st edn. published 1976).

Dawkins, R. (2007) *The god delusion*. (Black Swan edn.) London: Transworld Publishers (First published 2006).

Dennett, D.C. (1995) *Darwin's dangerous idea: evolution and the meanings of life.* London: Allen Lane The Penguin Press.

Findlay, A. (1965) *A hundred years of chemistry*. (3rd edn.) Revised by Williams, T.I. London Methuen & Co. Ltd. University Paperbacks.

Finke, R.A. (1995) 'Creative realism', in Smith, S.M., Ward, T.B., and Finke, R.A. (eds.) *The creative cognition approach*. Cambridge, Massachusetts: The MIT Press (Bradford Book), pp.303-326.

Fodor, J.A. (1983) *The modularity of mind*. Cambridge, Massachusetts: MIT Press.

Fodor, J.A. (2000) *The mind doesn't work that way: the scope and limits of computational psychology*. Cambridge, Massachusetts: MIT Press.

Frankish, K. (2005) *Consciousness*. Milton Keynes: The Open University.

Gaut, B.N. (2003) 'Creativity and imagination', in: Gaut, B.N. & Livingston, P. (eds.) *The creation of art: new essays in philosophical aesthetics*. Cambridge: Cambridge University Press, pp.148-173.

Ghiselin, B. (ed.) (1985) *The creative process: a symposium* (California Paperback edn.). Berkeley and Los Angeles, California: University of California Press. (First published 1952).

Hadamard, J. (1949) *The psychology of invention in the mathematical field*. Enlarged edn. Reprint, New York: Dover Publications, Inc., 1954 (1st edn. published 1945).

Harris, P.L. (2000) *The work of the imagination*. Oxford: Blackwell.

Hastings, N.A.J. and Peacock, J.B. (1974) *Statistical distributions*. New York: Halstead Press.

Hausman, C.R. (1984) *A discourse on novelty and creation*. 2nd pub. Albany NY: State University of New York Press. (First published 1975).

Jackson, H.J. (ed.) (1985) *Samuel Taylor Coleridge*. Oxford: Oxford University Press.

Kant, I, (1793) *Critique of judgement*. 2nd edn. Translated by Pluhar, W. Indianapolis: Hackett, 1987. (1st edn. published 1790).

Kaufman, G.D. (2004) *In the beginning … creativity*. Minneapolis: Fortress Press.

Koestler, A. (1964) *The act of creation*. London: Hutchinson.

Kronfeldner, M. E. (2007) *Darwinism, memes, and creativity: a critique of Darwinian analogical reasoning from nature to culture*. PhD, Universität Regensburg [Online]. Available at: http://epub.uni-regensburg.de/10562/ URN: urn:nbn:de:bvb:355-opus-8015. (Accessed: 25 September 2014).

Kronfeldner, M.E. (2009) 'Creativity naturalized', *The Philosophical Quarterly*, 59 (237), pp.577-592.

Kronfeldner, M.E. (2011) *Darwinian creativity and memetics*. Durham: Acumen Publishing Limited (Acumen Research Editions).

Lasswitz, K. (1901) 'The universal library' Translated by Ley, W. In: Fadiman, C. (ed.) *Fantasia mathematica*. Reprint, New York: Copernicus, 1997, pp.237-243. (*Fantasia mathematica* first published 1958).

Lewis, C.S. (1994) *The discarded image: an introduction to medieval and Renaissance literature*. Canto edn. Cambridge: Cambridge University Press. (1st edn. published 1964).

Lewis, D.K. (1986) *On the plurality of worlds*. Oxford: Blackwell.

Livingston, P. (2005) *Art and intention: a philosophical study*. Oxford: Oxford University Press.

Mill, J.S. (1874) 'Nature', in: *Three essays on religion*. 2nd edn. London: Longmans, Green, Reader, and Dyer, pp.3-65.

Mithen, S. (1996) *The prehistory of the mind*. London: Thames and Hudson.

Muir, K. (1984) 'Introduction', in Shakespeare, W. (1984) *Macbeth*. Edited by Muir, K. Walton-on-Thames: Thomas Nelson (Arden Shakespeare), pp.xiii-lxv.

Newton, I. (1676) Letter to Robert Hooke, 5 February. In: Turnbull (ed.) (1959), cited in *Oxford Dictionary of Quotations* 6th edn. (2004). Oxford: Oxford University Press. (Seiko electronic version).

Nichols, S. & Stich, S.P. (2003) *Mindreading: an integrated account of pretence, self-awareness, and understanding among other minds*. Oxford: Oxford University Press.

Nickles, T. (2003) 'Evolutionary models of innovation and the Meno problem', in Shavinina, L.V. (ed.) *International handbook on innovation*. Amsterdam: Elsevier, pp.54-78.

Noble, W. and Davidson, I. (1996) *Human evolution, language and mind: a psychological and archaeological inquiry*. Cambridge: Cambridge University Press.

Ortony, A. (ed.) (1979) *Metaphor and thought*. Cambridge: Cambridge University Press.

Patrick, C. (1955) *What is creative thinking*. Reprint, Nabu Public Domain Reprints, no date. (First published: no date).

Perkins, D.N. (1981) *The mind's best work*. Cambridge, Massachusetts: Harvard University Press.

Pinker, S. (1997) *How the mind works*. Allen Lane The Penguin Press.

Plantinga, A. (1976) ‘Actualism and possible worlds’, *Theoria*, 42. Reprint, in: Loux, M. J. (ed.) (2001) *Metaphysics: contemporary readings*, London and New York, Routledge, pp168-187.

Plato, (1956) *Protagoras and Meno*. Translated by Guthrie, W.K.C. Harmondsworth, Middx.: Penguin Books Ltd.

Poincaré, H. (1908) ‘Mathematical creation’. Reprint, excerpt from Poincaré H. *The foundations of science: science and hypothesis, the value of science, science and method*. Translated by Halsted, G.B., 1924, pp.383-394, in: Vernon, P.E. (ed.) *Creativity*. Harmondsworth, Middx: Penguin Education, 1970, pp.77-88.

Quine, W. V. (1961) ‘Two dogmas of empiricism’, in: *From a logical point of view*. 2nd edn. revised. Reprint, New York: Harper & Row, 1963, pp.20-46. (*From a logical point of view* first published 1953)

Quine, W. V. (1990) *Quiddities: an intermittently philosophical dictionary*. London: Penguin. (First published 1987).

Rignano, E. (1923) *The psychology of reasoning*. Translated by Holl, W.A. London: Kegan Paul, Trench, Trubner & Co., Ltd.

Roberts, R.M. (1989) *Serendipity: accidental discoveries in science*. New York: John Wiley & Sons, Inc.

Rothenberg, A. and Hausman, C.R. (eds.) (1976) *The creativity question*. Durham, NC: Duke University Press.

Ruvoldt, M. (2004) *The Italian Renaissance imagery of inspiration: metaphors of sex, sleep, and dreams*. Cambridge: Cambridge University Press.

Shakespeare, W. (1984) *Macbeth*. Edited by Muir, K. Walton-on-Thames: Thomas Nelson (Arden Shakespeare).

Simonton, D. K. (1999) *Origins of genius: Darwinian perspectives on creativity*. Oxford: Oxford University Press.

Sorensen, R. (2002) 'The art of the impossible', in: Gendler, T.S. and Hawthorne, J. (eds.) *Conceivability and possibility*. Oxford: Clarendon Press, pp.337-368.

Standish, R.K. (2011) *Theory of Nothing*. 2nd edn. Australia: Booksurge. Kindle edn. Available at: http://www.amazon.co.uk. (Downloaded: 24 September 2014).

Sternberg, R.J. and Lubart, T. I. (1999) 'The concept of creativity', in: Sternberg, R.J. *Handbook of creativity*. Cambridge: Cambridge University Press, pp.3-15.

Stringer, C. and Gamble, C. (1993) *In search of the Neanderthals: solving the puzzle of human origins*. London: Thames and Hudson.

Taylor, M.C. (2001) *The moment of complexity*. Chicago: The University of Chicago Press.

Tooby, J. and Cosmides, L. (1992). 'The psychological foundations of culture', in: Barkow, J.H., Cosmides, L. and Tooby, J. (eds.) *The adapted mind*. Oxford: Oxford University Press, pp.19–136.

Vedral, V. (2010) *Decoding reality: the universe as quantum information*. Oxford: Oxford University Press.

Vernon, P.E. (ed.) (1970) *Creativity*. Harmondsworth, Middx: Penguin Education.

Wallas, G. (1926) *The art of thought*. London: Jonathan Cape.

Ward, T.B., Smith, S.M., and Finke, R.A. (1999) 'Creative cognition' in: Sternberg, R.J. (ed.) *Handbook of creativity*. Cambridge: Cambridge University Press, pp.189-212.

Whitehead, A.N. (1967) *Adventure of ideas.* First Free Press paperback edn. New York: The Free Press. (1st edn. published 1933).

Whitehead, A.N. (1985) *Process and reality: an essay in cosmology.* Corrected edn. Edited by Griffin, D.R. and Sherburne, D.W. First Free Press paperback edn. (1st edn. published 1929).