# As We Build our World We Build our Minds:

## The Causal Role of Technology in the Development and Evolution of Human Psychological Traits

By Matt Gers

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

Victoria University of Wellington

(2011)

## Abstract

Do genes or environments have more of a role to play in the development of psychological traits? The nature versus nurture debate takes many forms and recent developmental systems arguments consider the roles of developmental resources to be inextricably linked. In this thesis I show that some elements of human culture, specifically technologies, play a privileged role in psychological development. Moreover, as we invent new technologies, we change the developmental environment for the present and subsequent generations, thereby causing evolution of the mind. I begin by outlining evidence, which shows that culture, and technology in particular, cause novel psychological traits to develop. Then I explain the evolutionary dynamics by which novel technologies and traits co-elaborate each other. The brain has evolved adaptations for plasticity and responds to environmental challenges in novel ways during development. I also show that brains often integrate with the material world, incorporating symbols, technologies and other artefacts as part of distributed information processing systems. Having demonstrated that technology has a causal role to play in cognitive development and function, I then move on to explain how we can distinguish among causal roles and thereby favour some causes over others in explanations of the development of traits. Beginning with Woodward's analysis of three dimensions of biological causation, I build a concept space and incorporate a fourth dimension of causation. This modified four-dimensional concept space of causal roles allows us to categorize and distinguish the causal role of genetic and non-genetic developmental resources. It turns out that, with respect to some questions or effects we are interested in, genes are important, and with respect to many other effects or developmental outcomes, cultural technologies are more privileged causes. I illustrate the use of this causal analysis tool by explaining the development of critical thinking skills. I conclude by arguing that the privileged role of technology in psychological development may help to explain two problems in human prehistory. First, it helps to explain why there was a lag between *Homo sapiens* becoming anatomically modern and only later becoming behaviourally

modern, and second, it helps to explain the mysterious extinction of the Neanderthals.

## Acknowledgments

All the following people and organizations have contributed in some way, whether motivational, intellectual or financial, to the completion of this research: Kim Sterelny, Ben Jeffares, Sondra Bacharach, Ken Perszyck, Stuart Brock, Nick Agar, Simon Keller, Barry Lewis, Philippa Hay, The Tertiary Education Commission, Victoria University of Wellington, The Victoria University of Wellington Post Graduate Students Association, Stacie Chalmers, and the graduate students of the Philosophy Department at Victoria University of Wellington 2008-2011, who are too numerous to name (for fear of omitting anyone!)

## Table of Contents

# List of Diagrams

# INTRODUCTION – The Moghul Emperor Akbar and the Nuclear Apocalypse

> Evolution has been transferred from the organism to the environment… in the products of art, and the recorded examples of noble lives, we have an environment which is at the same time the product of mental evolution, and affords the condition of the development of each individual mind to-day. (Morgan 1896)

Imagine that there is a nuclear apocalypse. All adult humans are killed and all technology as we know it is destroyed. This includes computers, the Internet, books, all representational media, all social technologies such as education systems and syllabi too. Somehow a few young children manage to survive the effects of the radiation alongside rats, cockroaches and the odd genetically modified crop. The rats and cockroaches go about their daily business pretty much as before the disaster. They scurry about, eat food, breed, perhaps solve the odd maze or two. But what about the human children, how do they behave? Or more particularly, how do they think?

Unfortunately for humanity, a hideous undertaking suggests an answer: the Moghul emperor Akbar (1542-1605) conducted a similar experiment (Catrou 1826; Ridley, 2003, p. 169). Akbar, who conceived of his monstrous experiment as progressive science, was motivated from a weariness of religious warfare and decided to determine whether humans were innately Hindu, Christian or Muslim. Several children were raised in silence and isolation, in a fortified castle, tended to only by deaf mutes[1]. Brought before the court at the age of twelve:

> …these children appeared before the emperor, to the surprise of every one, they were found incapable of expressing themselves in any language, or even of uttering any articulate sounds. They used only certain gestures to express their thoughts, and these were all the means which they possessed of conveying their ideas, or a sense of their wants. They were, indeed, so extremely shy, and, at the same time, of an aspect

[1] Harlow & Soumi (1971) have conducted similar experiments with monkeys and report similar results.

> and manners so uncouth and uncultivated, that it required great labour and perseverance to bring them under any discipline, and to enable them to acquire the proper use of their tongues, of which they had previously almost entirely denied themselves the exercise. (Catrou 1826, p. 117)

One final (fictional) scenario appears in Jean Auel's 'Clan of the Cave Bear'. Here the *Homo sapiens* child Ayla is raised by Neanderthals. As she grows up her human cunning and intelligence enable her to succeed in the Neanderthal clan. Though an entertaining and sweeping epic, this book's premise is gravely misconceived. Ayla in all probability would have grown up thinking and acting just like her Neanderthal caregivers. As I will demonstrate later in chapter eight, in all likelihood if the premise had been reversed and a Neanderthal child was raised by humans then the child would have grown up expressing human competence with symbols and a degree of abstract thought probably not achieved by 'wild-type' Neanderthals.

Obviously, neither Hinduism, nor Christianity, nor Islam is innate. But I will argue that virtually nothing is 'innate' in our common conception of the term. This conception being some approximation of invariantness and having an evolved history (see chapters two and seven for discussion). Furthermore, our usage of the concept 'innate' is confused. Brains develop in a rich context. A highly flexible system of neurons is causally structured by the environment. Indeed, as we will see in chapter three, the environment in many ways is a part of the mind, or at least the supportive crutch that allows the mind to function as it does.

In what follows I will be particularly focusing on human constructed environments. I will demonstrate how the worlds we build in turn build our minds. Processes of niche construction, technological innovation, the structuring of education systems and the ongoing integration of brains with the world have driven human psychological evolution, particularly over the last 40,000 years. With a particular focus on technology, I aim to show that not only does technology and educational structure explain the unparalleled success of *Homo sapiens*, but the environments we are creating right now, are having real-time effects on how our children think. Not only do Akbar's deprivation experiments (and a host of animal data) show that impoverished environments alter psychological phenotypes, but enrichment data

(from animals and humans) show that brains are highly adaptive and plastic, responding in novel ways to novel environments. The constraints on development are not as stringent as some nativists have argued and a wide range of cognitive phenotypes can be built by structuring in the environment. I note, however, that the argument has certain important caveats. Sometimes simple restructuring of the brain can have extreme consequences for development and subsequent cognitive ability.

I am interested in what humans have done, can do, or might be able to do. Human action is driven by psychology. I will be taking a broad approach to what counts as a psychological trait. All skills, knowledge, or processes, mediated (in part, but not necessarily exclusively) by the neural system, in particular those that are the subject matter of cognitive science or educational psychology, and which help to determine human action, will be of interest. This will include thought processes, algorithms for solving problems, databases of factual knowledge, and variations in perception, language, attention, memory, intelligence, emotion, and so on. All such things will count as psychological traits for the purposes of the present argument.

The sociocultural perspective on psychological development championed by Vygotsky (1978) claims that children develop cognitive structures in response to the use of cultural tools and language. The developmental cultural environment accordingly influences what cognitive devices the child acquires. I plan on assessing the evidence for this approach to cognitive development. I will try to establish in which domains, if any, culture determines psychology; and if particular cultural objects or technologies are the difference makers in the formation of some cognitive device rather than another. There is, furthermore, a debate over whether any causes of development carry information about, or represent, the outcomes of development. For example, do genes ever 'code' for traits? In chapter four I will outline these arguments, and deny that genes can be considered semantic information bearers or developmental codes. I will explain what it would take for anything to be a bona fide developmental code, and in chapter seven I suggest that almost nothing can satisfy these criteria. Cultural objects and technologies are unlikely to code for their causal effects in development.

All this throws up a further important question, which will become the focus of several later chapters. This is the main question my thesis addresses. If there are multiple, necessary, interacting causes of development, how are we to apportion causal responsibility in explanations of development? If technology is a cause of psychological development, and culture in general has a role to play, and other environmental features have their own influences, and genes obviously feature in these causal explanations, how do we explain which causes are important, and why? I argue that it is possible to understand complex development without committing to simple single factor determinist theories of causation on one hand, or being drawn into intractable systems approaches on the other.

To clarify, I am only interested in human psychological traits. Though some of the arguments may generalize to non-psychological traits, or to non-human traits, others will not. This is partly because of the special sort of organ of plasticity that the brain is. Once we see in chapter two what sort of thing we are dealing with when we examine the brain we will understand why my arguments may sometimes not apply to hearts, hair colour, or hormone levels.

## ARGUMENT PRECIS

### Part One – Cultural Technology and Psychological Traits

The argument is that human psychological development is influenced by developmental environments. Sometimes these developmental environments are partly constituted by human technologies. As technology changes over time, developmental environments change. This results in different psychological traits in different places and at different times. A key claim of my argument is that we can distinguish among causes in development and that sometimes human technologies are important or privileged causes of cognitive development. They, therefore, ought to feature prominently in developmental and evolutionary explanations.

The argument proceeds as follows: In chapter one I catalogue correlations between various cultural technologies and a range of psychological traits. In many instances the relationship is causal. This will show that culture can modulate, affect, attention, and some of the constituents of IQ, such as numerical cognition. Overall, I show that cultural technology is *a* cause of psychological traits.

In chapter two I outline an explanation of these effects. Culture causes psychological traits through experience-dependent neural plasticity. I discuss the neuroconstructivist approach to psychological development. Even taken in conjunction with nativist approaches, neuroconstructivist data suggests that it unlikely that anything is 'innate' in the common sense of the word, and that the construction of minds always depends on context. Rather than claiming environmental determinism, I am attempting to explain the mechanism behind the phenomena I illustrated in the first chapter.

Chapter three shifts attention from development to context and examines how constructivist brains integrate with, and depend upon, external media for their development and function. I discuss the notion that our brains integrate with and depend upon our bodies and external supports for thinking. Furthermore, sometimes these external supports constitute proper parts of our cognitive systems. This will demonstrate that it is not only developmentally that we are shaped by external factors, but that operationally minds require the integration of brains, bodies and the technological world.

To this point I will have outlined the observable effects of cultural technology on psychology, and three relationships between cultural technology and minds. The first effect is a causal developmental relation. I will show that technology has measurable effects on the outcome of development, causing some traits rather than others to emerge. Second, I demonstrate a functional dependence relation. In these cases technology plays an essential role in the operation of some human psychological capabilities. Finally, I argue for a constitutive cognitive relation. If this argument works for any given technology, the technology is to be properly seen as a component

of the agent's cognitive system. These arguments will lead into part two of the thesis where I examine kinds of causal roles in development in general and build a tool to catalogue the kinds of roles that cultural technologies (as compared to other factors) play in psychological development.

## Part Two – Untangling Causation in Development

In chapter four I introduce and explain a concept-space tool that we can use to distinguish among causal kinds in developmental systems. First, I introduce developmental systems theory. One of the key arguments that systems theorists have made is to reduce the role of genes to no more important than any other element in the developmental matrix. In response some argue that the informational content of genes is reason to privilege them in explanations about development. I argue that this is wrong. I then discuss the idea that genes are privileged, sometimes, in some contexts, and with respect to some questions. But they are privileged for the same reasons that any other aspect of the developmental matrix (including cultural technologies) may be privileged. Developmental causal privilege is a function of kind of causal role, not of information. By taking a causal analysis approach we can bring structure to developmental systems theory and to nature-nurture debates. The tool I introduce for analysing and representing causes does just this.

In chapter five I apply the tool generated in chapter four to several causal factors in development. I demonstrate that genes are neither a universally privileged cause, neither do they play a homogenous role in development. They are sometimes important, but so are some cultural technological causes of psychological traits. I emphasize that not all causes are equal, and that the tool introduced in chapter four can help us analyze whether causes are privileged, and why.

Chapter six details a case study that uses the tool to analyze the causes of one psychological trait, the ability to think critically. I explain why many of its causes are not relevant to causal models of this trait's development. I then construct a toy

example of a causal explanatory model for this trait and emphasize the importance of cultural technologies in its development. This move sheds light on the importance of the role of cultural technologies in causing the psychological traits we encountered back in chapter one.

The concept space introduced in chapter four can be used to understand some concepts in biology. I show in chapter seven that regions of the space and the causes therein help us understand the concept of innateness, and what it takes to be a code for development. This chapter summarizes the theoretical implications of the causal analysis tool.

### Part Three – Evolution and the Sapient Paradox

Finally, in chapter eight, I turn from development and causes to evolution. If psychological development in part depends on cultural technological environments, then psychological evolution will occur when such developmental environments change. Understanding this helps explain the 'sapient paradox' (Renfrew 2008). This is the question of why there is such a long temporal gap between the 'modern' human genome being established and the significant behavioural and cognitive take-off of the Holocene (the last ten thousand years). A related question is why the comparatively similar species *Homo sapiens* and *Homo neanderthalensis* took such radically different evolutionary trajectories leading up to the Holocene. The fates of these species could not have been more different. Neanderthal became extinct and *Homo sapiens* went on to dominate the planet. In this final chapter I explain how parts one and two of this thesis contribute to resolving these questions.

To reitirate, the overall thesis argument is that as we build our worlds, by inventing and propagating technological innovations, we are also shaping our minds through developmental processes influenced by technological innovation. This is paradigmatic of, and exemplifies, the theory of cognitive niche construction. Psychological development is a complex process, but we can demonstrate on principled grounds,

that technological culture plays an important, privileged, role in the emergence of some traits.

# PART ONE:

## Cultural Technology and Psychological Traits

# CHAPTER ONE – Technology is a Cause of Psychological Traits

## Introduction

In this opening chapter I will show that cultures, and in particular technologies, have causal influence over psychological traits in humans. After a brief discussion of what I take technology to be, I will catalogue several different domains of human psychology that are open to modulation by cultural context. I will describe variability in intelligence, mathematical cognition, affect and attention. I show that all of these are partly or largely constructed according to cultural technological context. My general strategy will be to provide evidence that there is variation for each domain across or between populations (this may be temporal variation across relatively short timescales), and then to follow this with examples of which technologies are implicated in producing some of the variation.

The reason for describing these phenomena is so that I can highlight the wide range of potentially important components of the developmental matrix with respect to human psychology. Once we understand the sorts of contexts that produce a range of cognitive phenotypes we can start to identify particular sorts of causes of human psychological traits. This will help us eventually to bring structure to the complexities of development and to answer questions about prehistoric human minds by correlating variation in cognitive phenotype with variation in environments.

There are many cultural influences on psychological development. For example, poor maternal nutrition causes low birth weight. This has been shown to lead to low cognitive scores and poor problem solving ability (Walker et al. 2007). These researchers also found that caregiver sensitivity, responsiveness to the child, and caregiver affect, all led to differences in cognitive outcomes. As did 'cognitive stimulation', which sometimes resulted from teaching mothers to develop children's skills. In this thesis however I am interested in technology. What role is technology playing in the development of psychological traits? What role has human constructed

technology played in modulating psychological development over historical and evolutionary time?

By cultural technology I mean anything that is invented and made by humans, that is open to innovation, and has a public dimension, in that the entity can be shared between individuals as well as modified, formalized and potentially documented by them using external media. I want to rule out things that are purely behaviours, unless those behaviours depend upon tools. Tools might be anything from lists on clay tablets, to stone-tipped spears, to personal computers, or a syllabus in formal physics education. If behaviour depends in part on tools then it is technological. I count words because these are open to innovation, invention, can exist as public representations and be shared among individuals. Clearly there can be paradigm and more marginal cases of technologies. Words are perhaps a marginal case, whereas an iphone is paradigm case.

Nickerson (2005) conceives of technology as the building of artifacts or procedures (tools) to help people accomplish their goals. I am happy to accept that procedures are technologies. Behaviours on their own are not, but if the behaviour depends on a procedure that can be invented, modified, shared, formalized, documented and so forth, then it is technological. Techniques will be technological if they are procedures or if they depend on tools.

## Intelligence and the Flynn Effect

The intelligence quotient (IQ) measures the age-adjusted ability of individuals to perform certain mental tasks. It is established that the average population IQ has been rising over the one hundred years since the tests were invented (Flynn 1987, 1999). This phenomenon is known as the 'Flynn Effect' and is well explicated in Flynn (2007). There is much debate surrounding what it is that IQ tests actually measure (see e.g. Steen 2009) but for my purposes it is sufficient that IQ measures something consistent. The fact that IQ has demonstrably and uncontroversially changed shows

that there has been a cognitive change. This cannot be genetic as the effects can be measured in a single generation.

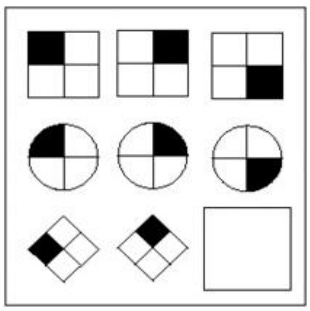
In industrialised nations the rises have measured 1.33 standard deviations in one generation. In rural Kenya the rises between 1984 and 1998 were even more dramatic (Daley et al. 2003). Various hypotheses have been put forward as possible explanations. These include: improved nutrition and health care, changes to family structure, improved schooling and increases in environmental complexity. Possible specific explanations for this 'Flynn effect' abound, but models that appeal to improved diet, genetic changes, or experience practicing the tests have been largely ruled out (Flynn, 2007; Johnson, 2005). For example, data from the Daley study suggests that health in Kenyan children was worse in 1998[2]. Also these researchers have controlled for methodological issues unlike several of the western studies. For example, the IQ tests given in 1984 and 1998 are unlike the Kenyan children's normal schooling. The same cannot be claimed in the west. The main remaining explanations are improved schooling and 'environmental complexity'.

We can measure different aspects of intelligence. These are classed as crystallized intelligence and fluid intelligence. Crystallized intelligence refers to the set of factual knowledge that an individual learns and includes such things such as vocabulary and times-tables. A person's education and experience determines what they learn and how they are able to apply this knowledge. Crystallized intelligence also increases with age (Toga & Thompson 2005). Fluid intelligence refers to more analytic reasoning and memory capacities and the processing speed of cognition. This declines with age.

Interestingly the observed IQ rises are not uniform across all IQ sub-tests. The rises occur largely in the Raven's Matrices and Similarities sub-tests (see figure 2). Changes in these tests have been shown to be up to two standard deviations in 50 years. These sub-tests assess abstract thought and fluid intelligence. The similarities sub-test

[2] These researchers don't draw this conclusion from their data, though it is clear that the incidence of significant anaemia increased over the study period. Anaemia in Africa is almost certainly due to hookworm and treatment-resistant malaria.

asks questions like, ‘in what way are dogs and rabbits alike?’ Figure 1 shows an example of Raven’s Matrices.



Firgure 1: A Raven's Matrices test item. The subject is required to select a figure that will complete the pattern.

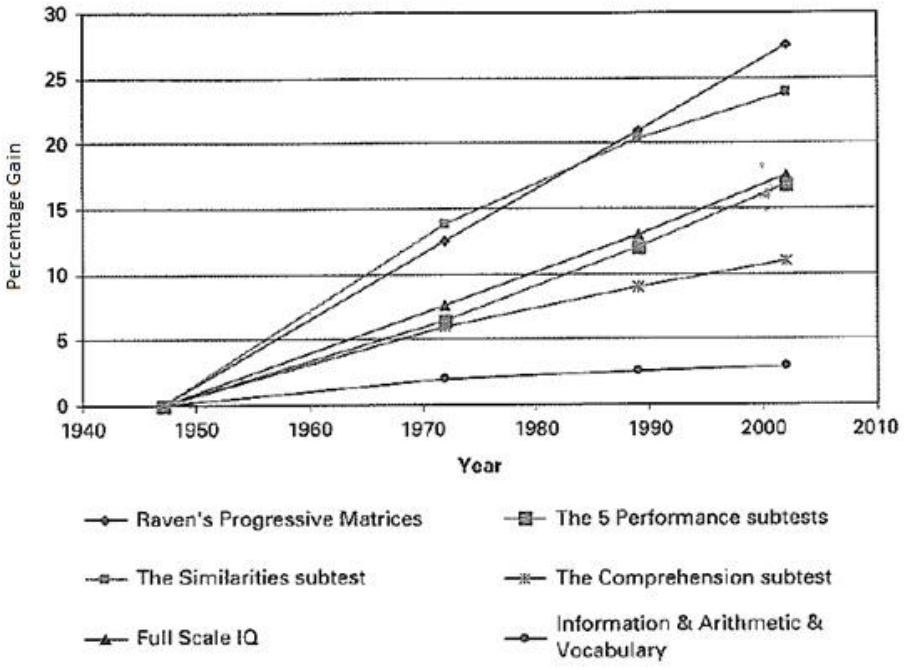



Figure 2: Percentage IQ Gains in the Twentieth Century. Note that the Raven's Matrices and Similarities sub-test items have shown dramatic change, whereas arithmetic and vocabulary are relatively static. (Adapted from Flynn 2007).

So, fluid intelligence has increased. Overall I am less interested in whether high IQ is a useful trait to possess. I am merely interested in the fact that there have been population-wide changes in psychological traits over brief periods of time. Let’s now look in more detail at two of the proposed causes of these IQ gains. First ‘improved

schooling', and second 'environmental complexity' (of which improved schooling may actually be an instance).

One explanatory model appeals to changes in formal education technologies (including syllabi and techniques for presenting material) over the twentieth century. This is the neurodevelopmental schooling hypothesis (Blair et al. 2005). This theory notes that in the first half of the twentieth century the wider availability of formal math education caused a population-wide IQ rise. And once math education was universally available, an increasing focus after the 1950s on providing fluid-intelligence-demanding problems to younger students continued the trend. For example, geometry, once the realm of university mathematics education, was taught to elementary students. Also, children taught arithmetic using traditional rote and algorithmic methods tend to use more rigid and less effective techniques to solve problems than those who learned socially and constructively, figuring out the properties of the decimal system through emphasis on concepts and manipulative materials. So it is proposed that this twentieth century increase in IQ was a two-phase process. Importantly, the skills required for these math tasks crossover with Raven's matrices and the Similarities IQ sub-tests. Not only do twentieth century school math text books include question items similar to Raven's Matrices (Blair et al. 2005), but the Raven's items test the ability to induce abstract relations and manage a large set of problem solving goals in working memory (Carpenter et al. 1990). All this suggests that changes to math education practices, especially when more and more abstract concepts are being taught to younger students, will drive IQ gains. The reason why literacy and numeracy scores have remained relatively static is because reading and arithmetic have been part of ordinary human society for a long time, whereas more abstract categorization of the world is a relatively new phenomenon and is associated with the more recent scientific worldview.

Science education is, therefore, likely to be another contributing factor. Just as experience with mathematical symbols can open up new conceptual possibilities, so experience with scientific concepts can open up new categories of thought and therefore new possibilities for action in the world. Flynn (2007) describes some

questions posed in the similarities sub-test. When asked about the relationship between a rabbit and a dog, students not educated with a scientific worldview state things like, 'a dog chases a rabbit.' The thought is concrete. But those scientifically educated may suggest that both animals are mammals. New possibilities for describing the world emerge. The important point, and a theme of this thesis, is that changes to pedagogical procedures, content, and method, change psychological traits in the developing generation.

Another proposed contributing cause of the Flynn effect is 'environmental complexity'. As with laboratory animal cognition (e.g. Lewis 2004, also chapter 2 of the present work), 'environmental complexity', loosely construed, has been suggested as a factor in the rising IQ scores seen globally. Johnson (2005) suggests that popular media has had a dramatic effect. He compares the cognitive demands of contemporary media with those of a few decades ago. For example, 'Dragnet,' a 1970s detective show, tended to follow a linear plot, always from the detectives' point of view. 'The Sopranos,' a more recent mafia series, has multiple plot strands, running through many episodes, with point of view changing from scene to scene. Similarly, the social network of 1980s show 'Dallas' involved only a handful of key characters, whereas the more recent show '24' requires the viewer to track the relationships of a few dozen. Johnson also argues that videogames have become fiendishly complex and maddeningly hard. Furthermore, by encouraging players to probe, hypothesize and re-probe, they are encouraging scientific thought. Johnson argues that hypertext and the Internet have changed the way we think, and have encouraged us to manipulate information, to act upon it rather than just let it wash over us.

It may be difficult to test relative overall complexity because as some things get more complex others become simplified. For example, there is no doubt that the visual symbolic complexity of our environment has increased during the twentieth century, however, the general trend in many parts of the world to a smaller family size has reduced the social complexity of particular aspects of people's lives (and conversely increased social complexity in other aspects). So we may see a variety of cognitive

effects, enhancements to visual symbolic processing, and possible decreases in cognitive aspects such as empathy (see the section titled 'The Affect Effect' below).

The argument is that it is our children's mental diet that has changed from 1900 to 2000, not their edible diet. In 1900 children might help with household chores, play stickball, read books or even undertake manual work. In 2000 they are creatively using text in instant messaging, following masses of baseball stats, entering virtual worlds where physical laws are broken, and following intricate plots on television series. As we will see in chapter two, the human brain is malleable and responds to what it experiences. Today we feed a more and more complex diet to our brains. Hence, argues Johnson, popular culture in part drives the Flynn Effect.

We see stepped stimuli throughout the twentieth century; the complexity of media, games, communication, and school syllabi has progressively increased. Of course higher IQs also thirst for more complex puzzles, so a pop culture ratchet begins. If this is true, we should see marked inter-cultural effects, and rural-urban effects too, because socio-cultural factors are not uniform in their effects on populations. Alexopoulos (1997) compared rural and urban IQs in Greece in 1979 and in 1997. He found a difference of 12 IQ points favouring urban Greeks in 1979. This had reduced to 4 points by 1997. The suggestion is that increased media penetration, transportation, and an overall rise in 'cultural level' in rural areas has driven this. This is entirely consistent with the drivers of the Flynn Effect that I have already described. Sternberg & Preiss (2005) quotes similar effects in the USA.

The Flynn Effect is not lost on those who monitor popular culture as critics either. In an essay about the 2010 film 'Inception' (which subsequently won four Oscars at the 2011 Academy Awards), Patrick Goldstein of the Los Angeles Times writes:

> 'Some critics have raved about the film's originality while others have mocked its excesses. If you were a young moviegoer, you loved the visually arresting puzzle-box thriller. But the older you got, the more likely you were to detest the film's run 'n' gun, dream-within-a-dream complexity.'

The idea is that the younger generation who have grown up on a diet of computer games, with their range of worlds and levels, more easily track the narrative density and clues of 'Inception' (Goldstein 2010). There is no doubt that a switch to visually based cognitive development in the late twentieth century is influencing cognition.

It is argued that intelligence is strongly heritable (e.g. Toga & Thompson, 2005). Therefore there is a paradox in postulating large environmentally-dependent changes in IQ over a couple of generations. But Dickens & Flynn (2001) resolve this paradox. A key part of the explanation follows Lewontin's (1995) arguments about gene-environment interactions. Lewontin notes that plant height might be largely genetically determined, but we could take a sack of corn and divide it in half. Plant half and fertilize it, and plant half and not fertilize it. The (significant) difference in mean height between the two populations would be the result solely of environment. So the question is this: is there an 'IQ nitrate' that has been in action in human environments in the last century? This would be one possible explanation.

Dickens & Flynn (2001) resolve this paradox further by explaining a model in which people's IQs are determined by genes and their environment, but in which environment is matched to their IQs. Basically this is the idea that people with slightly higher IQs may be rewarded for good test scores, or find intellectual pursuits more enjoyable, so they seek out environments like libraries and practice math problems in their spare time. This would lead to further intelligence gains and so on. The idea is perhaps simply elucidated by imagining two genetic variants, x and y, and a range of environments A through E. We can plot a table of the hypothetical IQ scores of individuals with each genotype in each environment (figure 3):

| Environment | A | B | C | D | E |
|---|---|---|---|---|---|
| Genotype | X / Y | X / Y | X / Y | X / Y | X / Y |
| IQ | 90 / 100 | 90 / 105 | 90 / 105 | 95 / 110 | 100 / 130 |

Figure 3. Hypothetical Gene-Environment Mappings

But importantly, possessing genotype Y tends to cause people to select environment E. It is hard to see how IQ could be described as 'innate' given these interactions.

## Mathematical Cognition

Having suggested some pedagogic and technological influences on IQ, I will now explore mathematical cognition. Language has an effect on human cognitive development. It enables us to do things with our minds. Speakers of Piraha (which lacks consistent number words entirely) although able to match quantities and therefore have an appreciation of the concept of exact quantity, are unable to remember and compare cardinalities across space and time and therefore lack basic arithmetic skill. It also seems that Piraha speakers do not represent exact cardinality mentally (Frank et al. 2008, Pica et al. 2004, Everett 2005). This phenomenon can be explained once we understand a little about the mathematical competences of the primates.

Numerical cognition in primates and humans exhibits some similarities but many striking differences. It appears that there are at least three systems involved in number concepts. A low-numerical-value exact system enabling comparison of numbers up to about three or four, which primates clearly possess as they can make selections preferring two to one, and three to two food items (Hauser et al. 2000). There is a large number approximate system, which permits comparison between larger numerical concepts. But the resolution of this system is limited and the larger number of objects must be approximately 1.5 times the size of the smaller (more for infants, less for adult humans). This is the Weber limit on performance. Human adults possess a Weber limit at a ratio of about 1.15 to 1 (Lipton & Spelke 2004). Discriminability depends on this ratio. Rhesus monkeys were unable to discriminate between 3 and 8 on a choice task. Cotton-top tamarins can represent number cross-modally but not precisely (Hauser et al. 2003). So it appears that primates possess but lack significant skills with this device. The third system is the large number exact system. This is necessary for exact arithmetic operations with large numbers, for example calculating the discrete sum of 12+19. It appears that this system is not only

language-dependent but also depends on the language learned. As noted above, speakers of Piraha (Gordon 2004) and Munduruku (Pica et al. 2004), which lack number words, have no real skills with large numerical manipulation. Here we see an example of rich language, or 'environmental complexity' in the more general Flynn Effect terminology, enabling a psychological trait, number sense. There is no large-number exact system without environmental support.

Speech, and more particularly, the specific language spoken can play an essential role in the organization of higher functions. Language, as I have noted already, is a technology, albeit a less paradigm case. We can craft it and employ it as a tool. Dennett (2000) is explicit about the nature of language as a tool. He uses the phrase 'tools for thinking' when referring to words.

Let's move on from language to symbols, particularly representation using the number line. When asked to perform mathematical manipulations that involve employing negative numbers children perform very badly and respond with statements such as, 'you can't do it'. Their conceptual appreciation of negative quantities depends on either exploring the physical number line, investigating the minus sign, or perhaps both. Without encountering the number line as a model, negative quantities are unimaginable. The problem, 'what are four apples minus ten apples' has no solution (Vlassis 2008).

If asked, 'can you write "no cats"?' a pre-reading child says, 'you can't write that cause there are no cats' (Olson 2005, p. 63). Concepts that relate to existing real-world objects such as 'cat' can develop without the symbol or word for 'cat', but the concept 'negative three' can't exist without the number line, minus symbol, or both. Saying, 'the opposite of three', or 'not three', fails to capture the concept, as the solution here intuitively seems to be 'zero', or at least 'nothing'. In normal human ontogeny, the acquisition of the concept of negative numbers requires public symbols. The concept of debt is not enough to scaffold the development of concepts of negative quantities because a debt itself is always a positive sum.

Symbols that don't relate directly to the material world can permit entirely new relations or concepts to be understood. Indeed, symbol-trained primates can learn to manipulate second order relations of sameness and difference through symbol identification (Thompson, Oden, Boysen, 1997) whereas wild primates never do. Chimps can be trained to recognize relations of sameness and difference and tag pairs such as cup/cup or cup/shoe with the appropriate arbitrary symbol. Important is the next step of the experiment where tag-trained chimps, and only tag-trained chimps, could then repeat the task with higher-order sameness and difference relations using pairs such as cup-cup/cup-shoe. The idea is that the tag training allowed the chimps to reduce the higher-order problem to a descriptively less complex scene by mentally substituting each pair with a sameness or difference symbol. They could then take this new pair of just two objects and repeat the process. Humans of course make many such substitutions every day. Something as simple as putting all my lollies in one bag and all yours into another bag reduces the descriptive complexity of a task. Labelling the world with tags, containers or categories increases our computational opportunities. By learning to label the world a host of previously imperceptible relations become apparent and useful to us.

The notation system learned influences the difficulty of mathematical operations. Imagine trying to do long division with Roman Numerals. It is not impossible, but is considerably more difficult. The Arabic system is more suited to the problem. At one grain of analysis the function (coming up with the right answer) is the same. But at one grain of analysis the flight of birds and of jet aircraft is the same. Yet there is something importantly different and interesting about birds as compared to jet aircraft. Marr (1982) distinguishes between the functional and the algorithm level of a process. A process may count as 'doing arithmetic', in the sense of mapping two numbers to one sum. This does not depend on the system employed. However, at a different level of analysis we are choosing particular representations to input and output, and an algorithm by which the process is achieved. This second level specifies 'how' we are performing the calculation. The choice of an algorithm is constrained by what it has to do (say, long division) and the hardware upon which it must run (some combination of symbols and human brains). The differences in how a cognitive

manipulation is performed are real and important differences in many instances. Some symbols open up new possibilities for cognitive manipulations that involve physically shifting external symbols on external media. The cognitive process of calculation then supervenes on this system of brain, pencil and symbols. Once new rules or concepts have been created then new cognitive possibilities open up. I will discuss these issues further in chapter three.

Language and symbols have important roles to play in mathematical capability. Different linguistic and symbolic environments can lead to different mathematical capacities. I want to move now away from the general area of intelligence and mathematical cognition, to look more closely at a completely different domain of human psychology, that being affect and emotion.

## The Affect Effect

Affect is a term used in psychology to recognize someone's emotional state or the way they are feeling. It also denotes the outward appearance of such states. I now describe research that suggests there has been recent cultural modulation of affect. I consider empathy and aggression.

Eisenberg (2002) defines empathy as, 'an affective response that stems from the apprehension or comprehension of another's emotional state or condition, and that is similar to what the other person is feeling or would be expected to feel' (p. 135). Research demonstrates that empathy has changed over recent decades. In a meta-analysis of 72 studies using the Interpersonal Reactivity Index,[3] Konrath et al. (2011) show that college students in 2009 score significantly lower than those in 1979 on tests of empathic concern and perspective taking. Furthermore, the drop in empathy was most marked since 2000. In further analysis the same authors found that

[3] The Interpersonal Reactivity Index is a 28-item scale that gauges aspects of interpersonal sensitivity including: Empathic concern (feelings of sympathy for others' misfortunes), perspective taking (how people imagine others' points of view), fantasy (how people identify with fictional characters in books or movies), and personal distress (how people feel when they see the misfortunes of others).

perceptions of people's kindness and helpfulness have changed throughout the same time period.

It would be interesting to know if this phenomena exhibits cross-cultural differences. Are there particular cultural factors that are driving the effect? Konrath et al. comment on the cultural and social composition of the college student population over time and conclude that it has not changed markedly in the study period. But what about non-college students? And are there rural-urban differences as we saw with the case study on IQ shifts? This sort of information would be useful to know. Konrath et al. attribute these changes to synchronus social and cultural changes. These include an increasing emphasis on the self, a preponderance of violent, horrifying media and the growth of online social media.

> 'With so much time spent interacting with others online and not in reality, interpersonal dynamics like empathy might certainly be altered. For example, perhaps it is easier to establish friends and relationships online, but these skills might not translate into smooth social relations in real life....' (Konrath et al. 2011, p. 9).

It may also be that the demands of college life have increased such that students don't have sufficient time to worry about others.

This large study merely shows a broad global effect of cultural context shifts on affect, specifically empathy. However, more refined research suggests that particular elements of culture may be contributing to the shift in empathy. Technology may play a role in this shift over time. It could be that the ease and speed of communication technology leads people to become more easily frustrated or bored. Perhaps the content of media may desensitize people to the pain of others or increase feelings of personal threat (Konrath et al. 2011). Whatever the combination of causes, I will now discuss one example. Research shows that violent computer games are causally implicated in aggressive behaviour.

### Computer Games and Affect

Swing and Anderson (2008) review thirty-two independent samples where the effects of violent games have been studied. They conclude that there is unequivocal evidence that short-term exposure to violent games produces immediate increases in aggression and that repeated, long-term exposure increases aggression across the lifespan. They further argue that poor methodology has *under*estimated this effect in previous studies. Some studies demonstrate that the correlation is causal. Violent responses still take provocation, and violent videogame exposure is neither necessary nor sufficient for aggressive responses, but rather is probabilistically causal in the same way that smoking causes lung cancer (clearly other factors are part of the causal matrix). The effect is supposedly more important than exposure to violence on television because the agent can rehearse the whole action sequence from threat appraisal to violent act and is intimately identifying with the aggressor, usually from a first-person point of view.

Further analysis of videogames by Gee (2008) sheds light on why games are so good at teaching participants. Indeed, Gee argues that good videogames (by which he means entertaining and popular ones) illuminate how we learn. It is not a trivial feat to master something as complex as some contemporary videogames. So the game designers have built into the game features like safe practice arenas, just-in-time information cues and smart virtual artifacts. Game levels advance in difficulty so as to be on the edge of the player's present capabilities. All these techniques conform to what is known about optimizing learning. The most complex yet learnable games even shed light on learning strategies not yet used in education. Games are powerful educators. They are also popular because of their continuous reward structures. This satisfies the brain's dopaminergic system's yearning for payoffs.

With violent videogames we have a situation where real world players are in command of multiple identities, real and virtual. The identities can influence each other. This is indeed the case with many kinds of games, from role-playing with war paints and sticks to immersive virtual reality. Technology has both intended and

unintended effects (Sproull & Kiesler 1992a, 1992b). The intended effect of violent videogames is entertainment. An unintended effect is increasing the player's disposition to aggression. This effect is likely to be one (among many) specific contributing factors to the empathy changes noted by Konrath et al. over the last few decades. It is important to note that I am not claiming that this is the only cause of empathy changes in the last decades. But I am suggesting that effects like this one will conspire to produce the effects measured in populations over time.

## Attention

Cognitive science has generally concluded that attention has a limited capacity (Kahneman 1973). This is the basis behind laws that prohibit cell phone use while driving. Whether hand-held or hands-free, studies have shown that brake reaction time and object detection are impaired and accident rates rise (Watson & Strayer 2010). However, given the arguments so far, we should not conclude that this capacity for attention is immutable. In fact it may be that different technological contexts cause differences in attentional capacity. Individuals from different developmental environments may perform differently on split-attention demanding tasks. I will present data below regarding computer games and attention that suggest that varying use of computer games may have some correlation to attentional capacity, but the correlation unfortunately is not conclusively demonstrated to be causal. However we will see that the argument for differing attentional capacities as a result of differing developmental contexts remains plausible.

Watson & Strayer (2010) demonstrate that multi-tasking attentional differences do exist across populations. They found that virtually everyone tested on an experimental protocol that simulated cell-phone-while-driving showed decreases in performance. However, there was a small population of about 2.5% of subjects who showed no impairment at all[4]. These same 2.5% scored consistently in the top

[4] The 2.5% was 15 fold higher than a 100,000 trial Monte Carlo simulation run on the data.

quartile across a range of driving-related and auditory operation span tasks. This runs against the prevailing cognitive science wisdom that attention is always degraded in multi-tasking activities.

However, there is a general trend in cognitive science to search for similarities rather than differences across a population. Often the task of a research program is to deduce how a particular cognitive task is effected *in human beings* (perhaps as opposed to, say, chimpanzees). However, arguments in the next chapter defending the plasticity of brains ought to cast doubt on this approach to cognitive investigation. Neural circuits may operate differently in different individuals. Questions perhaps should be framed as, what *proportion of individuals* express such and such a way of cognizing?

Watson & Strayer identify those whose attention does not seem to be degraded while briefly multi-tasking. But they also argue that Parasuraman & Greenwood (2007) have found evidence for another unusual sub-population. The 'odd man out' in their experiment was an individual who could sustain attention on a single task without exhibiting vigilance fatigue. It may be that super multi-taskers and the odd man out are opposite ends of a phenotype spectrum (those good at focusing on many things at once, and those good at focusing on a single thing for a long time). The interesting question is what is the cause of these trait differences? Many potential causes of attention differences across a population exist. I will now suggest that computer game use may be one of them.

### Computer Games and Attention

Dye et al. (2009) note that action video game players of all ages have enhanced attentional skills. This motivates the question of whether it is the games that cause this trait or not. Feng et al. (2007) found that ten hours of playing action games was enough to increase participants' useful field of view on a visual attention task. Tahiroglu et al. (2009) found that use of computers increases attention in the short-

term as measured by the Stroop TBAG test[5]. However, this was only for those for whom the computer game was novel. Participants for whom the computer game was a daily habit did not show improved attention. It may be that novel games increase frontal and prefrontal activation and executive functions such as attention. Overall, computer games seem to have some effect on short-term cognition but exactly what is unclear. Tahiroglu's study also concurs with Chan & Rabinowitz (2006) who found that more than one hour per day use is correlated with inattention. Barlett et al. (2009) argue that research into computer games and cognitive development suggests that games can enhance selective attention, concentration, tracking and useful field of view. Considering that seasoned players seem not to show the effects it may be the aspect of novelty rather than a particular game or games that is important. Either way, using computer games seems to have some effect on attention.

Again, I am not attempting to suggest that computer games are the sole cause of attentional variations in contemporary populations. I am merely suggesting that such technologies may be one part of a *web of causes* that are driving recent, rapid, trait variation and change across many domains of human psychology. Later, in chapter four, I will begin outlining a tool for analysing causal roles, which helps us untangle this causal web.

## Pedagogical Tools

My final illustration of the effect of technology on psychological trait development for this chapter is the case of pedagogical tools. I will outline remedial reading software and diagram use for explaining concepts.

Mioduser et al. (2000) explain that reading partly requires the 'natural development' of a suite of cognitive tools, and partly the 'formal acquisition' of a set of abilities. These researchers undertook an experimental study to see if features of computer-

[5] The Stroop TBAG test has five subtests that sequentially involve reading colour words that are printed in black, reading coloured colour words, naming coloured circles, naming the colour of normal words, and naming the colour of coloured words where there are incongruities of colour and meaning.

based learning materials could enhance the learning of early reading skills in children at high-risk of learning disabilities. Improvement on three tests was significantly higher in the computer group than control groups.

> 'Children seem to benefit from computer-based work not only at the specific skills level but also, as a result of their improvement in academic achievement, in terms of motivational and self-confidence levels,' (p. 58).

They attribute this result to the embodiment effects of the computer. The child can press the screen, hear the sounds, and see the words. This active, embodied, engagement seems to enhance learning. There is immediate feedback and concrete manipulation of language. Of course, the acquisition of reading skills is highly individualized. Close observation of the child's performance is followed by mindful pedagogic decision-making. But with an algorithm for individuation the computer can perform this role. Obviously this depends initially on the programmer's intention, but technology and teacher's intention can interact positively. The computer software, however, is clearly a part of the cause of the child's reading skills.

Another technology that is causally implicated in attaining cognitive skills is the use of diagrams. Overall, using schematics seems to enhance cognitive performance. When teaching cell biology, diagrams scaffold what we can learn and understand. Scheiter et al. (2009) found that a realistic dynamic visualisation when compared to a schematic one was less effective in teaching the concepts of cell replication. Only students who watched the realistic visualization twice (and not the schematic one) had lower learning outcomes. So schematic representations enhance learning about real natural processes.

As the complexity of our visual environment increases due to information and communication technology, we can expect our skill with schemas, diagrams and other abstract representations to increase along with understanding of the concepts represented by these schemas.

## Summary

In this chapter I have illustrated that some technologies play a causal role in psychological development. I also emphasized variation in psychological traits due to variation in the technological exposure of individuals across place and time. Sometimes there are qualitative and dramatic differences in psychological traits, such as when comparing whether number words are present or absent. Sometimes the differences are of quantity or efficiency, such as with variations in action computer game use. I have shown that these effects cut across many domains of human psychology, from intelligence to affect.

These examples underscore the importance of active engagement with the material and symbolic world when learning various concepts. It is also clear that technology can have unintended as well as intended consequences for cognition. This is demonstrated by the unintended increases in aggression derived from violent computer game play. Also, cultural change can drive population-wide changes in our psychological profiles. The Flynn Effect shows this. And finally, individuals within a population can have different developmental experiences and therefore different psychological traits.

I also emphasized the interplay between technology and education. Take Mioduser's literacy software for example. The computer is a designed product which requires syllabus planning by educators and face-to-face modulation via teachers. Cultural technological causes of many cognitive traits are probably often hybrid in this way. Education structures themselves, however, are cultural technologies (schools, classes, timetables, syllabi, pens, paper, calculators, tests, the internet, written feedback, and so on).

Finally, many of the cultural technological influences on cognition (and there are a great many more than those I have described so far) are examples of humans inventing technologies and then perpetuating those technologies. As we build our worlds in this way we are changing not just the operative context of the present

generation, but also the developmental context of present and future people. This is a key phenomenon that I will return to later when I discuss niche construction and evolution in chapter three. Context is something that can be inherited.

## Key Claims

The message that I am trying to convey is that there are a variety of causal relationships between technology and psychological traits. Technological innovation has the potential to lead to rapid shifts in the psychology of populations, and sometimes there are important differences in technological exposure between individuals. The nature and significance of these causal relationships are varied and will need untangling at some point. This is a project that I begin in chapter four. In the next chapter, I explore the neural shifts underpinning the psychological effects described so far.

# CHAPTER TWO – The Neural Mechanism and Development

## Introduction

I have just surveyed the influence that human technological culture has on our psychological development. The fact that technological culture influences our psychological phenotypes accords with the interactionist consensus in developmental biology (e.g. Sterelny & Griffiths 1999, Pennington et al. 2009, Levy 2004), which asserts that both genes and environments are important in developmental processes. It is not merely nature or nurture that is responsible for who we are but it is nature 'via nurture' as Ridley puts it (2003). In the case of technological culture it is nature via technological culture (and other assorted determinants). But what is our nature when it comes to cognition? What mechanisms underpin the culture-psychology interactions I have described in chapter one?

Different positions on the relative contribution of genes and environment exist. On one hand there is a stongly nativist school of thought. Such Evolutionary Psychologists claim that the majority of psychological development is determined by genetic influences. They further argue that these genes and gene networks were selected deep in the Pleistocene (1.7 million to 10,000 years ago) because they tended to lead to the development of domain-specific specialized cognitive modules, which solved the adaptive problems of that era. Evolutionary Psychology emphasizes the *adapted* nature of our minds. At the other end of the spectrum are various constructivist approaches, which see development as highly contextualised. Traits emerge as a flexible, general-purpose brain encounters the world. In this chapter I describe examples which indicate that the latter is more likely. Given that this position largely denies innateness, I will begin this chapter with a few paragraphs on what innateness is supposed to be and end the chapter with some reflection on this concept.

The structure of this chapter is as follows: First I make brief comments about the notion of innateness. I will then describe the positive case for a constructivist view of development. What emerges will be a picture of neural development that explains very clearly why we see the sort of culturally mediated psychological trait outcomes that I discussed in chapter one. I will not be denying adaptation. I will not be denying the central role of genes in development. What I am trying to emphasize is that plasticity of trait development is important, that sometimes causes other than genes are important. Developmental plasticity must be constrained by the characteristics of the developing system, but we will see that there is no reason to believe that there are many invariant cognitive structures, or human psychological universals. After outlining the case for constructivism I will then raise two important problems for Evolutionary Psychology. It may seem obvious but psychological traits will be determined in part by evolutionary history, and in part by individual experience.

Over the next two chapters we will encounter three relationships between technological culture and minds. First, in this chapter I will explain how technological culture can modulate neural development. Different cultural and technological environments cause the development of different neural structures in brains. Necessarily, much of the evidence supporting this comes from studies of animal brains. In the next chapter I will explain scaffolding and extending of minds. Sometimes technological culture can act as a support or scaffold, which enables minds to function in particular ways. And sometimes technological culture should properly be seen as a part of the rich information processing system that spans brain, body and world. In these cases cognition can be said to extend beyond the body.

## Innateness

There is a folk concept, which tries to capture the ‘nature’ arm of the interactionist consensus. This is the concept of innateness. Many important debates have been framed by an innate/not-innate distinction. For example the IQ debate that I introduced in chapter one is often framed as, ‘how much of intelligence is innate?’ The concept of innateness is also a part of folk wisdom. Some traits of an organism

are characterised as part of its inborn nature and other traits are not. In the debate over cognition, nativists (notably Evolutionary Psychologists), propose that there is a core set of innate representations. Proponents of 'nurture' or the acquisition of traits argue that complex representations are constructed only through experience. It is interesting to note that both sides of this debate classify traits by their causes. Innate traits have internal causes of a particular kind and acquired behaviours come from the external (psychological traits that arise from embodied experience are a bit of a grey zone). I will have more to say about this in chapter seven after I have introduced a model for cataloguing causal kinds. For now I merely want to massage scepticism about the whole idea of innate traits.

The innate/non-innate debate is complex. For example, high IQ can be caused by a particular genotype, FADS2, but only in children who are breastfed (Capsi et al. 2007). Is this an internal or external cause? Is the high IQ innate? Is phenylketonuria innate or acquired, given that it is in a sense 100% genetic (you only get it with a particular genetic mutation) *and* it is 100% environmental (you only get it if there is phenylalanine in your diet)? It is far from clear in what sense such traits are supposed to be 'innate'.

It might be argued that some traits are genuinely innate. For example, newborn infants smile in response to adult faces in the first couple of months of life. This might be considered 'innate'. But shortly I will give examples where species-typical behaviour at birth is easily modified by simple pre-natal interventions. What is it about infant smiles or their causes that lead us to consider them innate? I will show that our brains are not hard-wired or prescribed ahead of development. If cortex is not innate in this sense then how can behaviour be? These considerations cast doubt on the whole concept of innateness. After I have described constructivist development I will return to this idea of innateness. At least some authors agree that the folk concept is confused and perhaps empty. I think we can salvage a related concept to the folk concept, however, and I will explain an 'internal, stable, heritable, privileged causes' notion in chapter seven.

## Neuroconstructivism

Development is the progressive elaboration of increasingly complex structures. All humans undergo development beginning as small and simple organisms and becoming larger and more complex across time (Stiles 2009). Neuroconstructivism is the view that neural structures elaborate according to contextual influences rather than being genetically pre-specified. Here is an example illustrating the development of auditory localization in barn owls. Given the conservative nature of evolutionary developmental mechanisms, this example is likely to be relevant to humans too.

If juvenile owls are fitted with a visual prism that shifts their field of view they can adapt to a twenty-three degree shift in about seven weeks and maintain accurate orienting behaviour to sounds (Shultz, Mysore & Quartz 2007). Removal of prisms results in a slow return to normal orienting behaviour. Adult owls, however, adapt poorly to prisms. But perhaps most interestingly, juveniles returned to normal and subsequently given prisms again as adults do show adaptability! This suggests that the intervention as juveniles has altered their cognitive architecture. The cognitive systems underpinning orienting behaviour are flexible and their development can be modified by interaction with external artifacts.

The explanation of this phenomenon lies in the way the auditory localization neural networks grow in barn owls. In order to orient accurately there must be a smooth mapping between the auditory input and the degree to which the owl turns. In normal circumstances there is a map from signals at some angle of degrees to an output of orienting the same number of degrees as the input. In this normal case we have the situation (seen in figure 4) where the input (i) maps to the output (o) such that $i_1$ maps to $o_1$ and $i_2$ maps to $o_2$ and so on. However, once this system has developed then prisms are applied to the owls. The prisms cause misalignment of the auditory input to orienting output transformation by 23-degrees. To correct for this, the owl's neural network needs to align input (i) with output (o + 23-degrees). Figure 4 represents this by illustrating a map from $i_1$ to $o_2$, $i_2$ to $o_3$ and so on. The owls can

learn to shift this mapping over time. But for this to be successful additional neural resources must be committed. This is so that $i_n$ can map to $o_{n+1}$. By adding units to the network the owls introduce the extra degrees of freedom in the input-output transformations required for orienting behaviour. When the prism is fitted the correct output is no longer the originally corresponding one. By growing the network and adding this unit the birds compensate over time. This explains why the juveniles returned to normal then given prisms again as adults could cope. When returned to normal they shift back to the $i_1$ maps to $o_1$ and $i_2$ maps to $o_2$ correspondence. But the unit $o_6$ remains in place, and can be employed in adaptation as adults again.

The novel context caused juveniles to commandeer uncommitted brain tissue and add extra units to their standard neural network. The new network is qualitatively different from the standard adult one because it can cope with distortion of visual input. It now has provision for a range of distortions to visual input and can shift between them. The new device is flexible, the standard one is not. It would be interesting to know whether juveniles fitted with prisms from birth were able to switch back if the prisms were removed as adults. I suspect not because their standard network would merely be the same as the original standard network, just shifted 23 degrees without provision for flexibility. This case exemplifies genuine construction of contextually relevant cognitive tools due to experience.

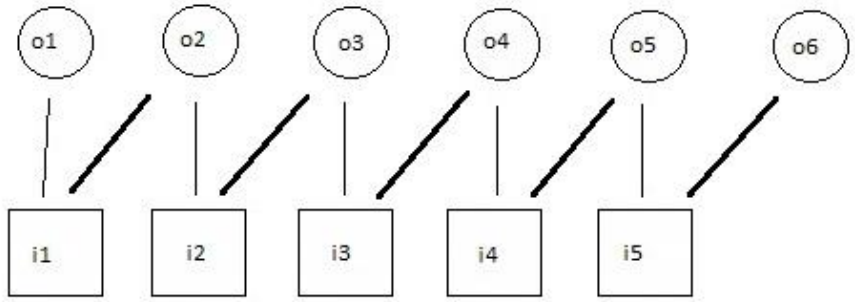



Figure 4: Constructing Barn Owl Auditory Networks. The row of boxes represents the inputs to a learning network. Initially the inputs are transformed into outputs in corresponding fashion, i1 to o1, i2 to o2, etc. When the barn owls are fitted with prisms to distort their visual field, the appropriate response to auditory input shifts. The adaptive response is for the owls to now output o2 for i1, o3 for i2, etc. This requires that the network grow and construct an additional unit, o6. Juvenile birds can achieve this over time because of neural plasticity mechanisms. With both mappings in place (thin lines for the original transformations and bold lines for the novel ones) the birds can switch between the appropriate responses as required. The new network is qualitatively different to the old one.

A standard view of learning is that networks learn by modifying connection strengths between nodes (e.g. brain cells). In brains this means modifying the thresholds at which individual neural junctions (synapses) will activate. This can and does enable learning but examples like the barn owl one suggest that we don't only learn by modifying connection strengths. Synaptic weight change may explain short-timescale plasticity, but other architectural changes occur on other time-scales. In fact if the only learning undertaken is the modification of synaptic thresholds then networks will be limited in what they can learn, because changing weightings that are already learned has the potential to destroy previous learning. Once these networks attain a certain state, then there are things that it will be impossible to learn. On the other hand networks with the ability to grow, or commandeer new structure are capable of solving any learning problem that can be solved, by any algorithm at all. Such constructivist networks are highly flexible and have true computational generality. 'The burden of innate knowledge is relaxed' (Shultz, Mysore & Quartz 2007, p. 71). Genes may have been selected that determine the initial mapping, but we cannot appeal to genetic adaptationism to explain the novel input-output map, or the flexible nature of the mature networks in adults-who-had-prisms-as-juveniles. The prism itself, as a novel object in the environment is a cause of the novel networks.

The emergence of the novel networks was enabled by inherent flexibility in the developing system. I argue that it is right to claim the flexibility is an evolutionary adaptation, but not the network itself.

Mine is a constructivist account of development. Developmental psychologists Piaget (1955), and Vygotsky (1978), whose theories drove developmental psychology for many decades held constructivist views. More recently neuroscientists Quartz and Sejnowski (1997) and Mareschal et al. (2007) have pushed similar lines. They call this account neuroconstructivism. A constructivist approach does not exclude evolutionary explanation of the human mind. In fact it depends on at least some tenets of an evolutionary account. A rich understanding of development is going to shape the degree to which we find Evolutionary Psychology on one hand, or neuroconstructivism on the other hand plausible. I will show how cognition is constructed contextually at every level of analysis. I'll give several examples of neuroscientific and cognitive research to support this view.

The neuroconstructivist picture examines development at the level of single neurons, brain systems and whole organisms. Neuroconstructivism notes the following. First development occurs in a rich context. Genes are active in a context of other genes; cells in a context of other cells; neural networks in a context of other networks; cognitive devices in a context of other cognitive devices; brains in a context of bodies; and bodies in a social, physical and technological environment.

The basic tenet of neuroconstructivism is that 'development is fundamentally constructivist in nature' (Mareschal et al. 2007, p. 5). There is a progressive increase in the complexity of representations. This results from re-wiring neurons and novel neural connections. New cognitive abilities form as a foundation of simple ones emerges and is modified. Over developmental time there is creation of genuinely new cognitive abilities as well as the better use of pre-existing ones. In the case of cognition, the structures are increasingly complex representational forms. Initially there exist only simple representations, and over time through mechanisms of competition and cooperation between elements, richer representations emerge,

which in turn allow proactive exploration of problem space and progressive specialization of neural and cognitive devices. The representations that emerge are partial and they are distributed across many functional circuits. Neuroconstructivism differs from dynamical systems in that there is a focus on representations, and it differs from classical connectionism in that the architecture of the system is not static, the system continues to build itself.

The emphasis here is on a few relatively simple processes that are repeated across different levels of description. Cooperation, competition and progressive change in structure over time occur at the level of neurons, neural nets, and cognitive devices. Representations do not emerge in isolation. The development of cognition is primed to be flexible. This is demonstrably the case as we will see in the examples below. To think that the same suite of cognitive devices will emerge across a wide range of contexts is a mistake. As environments change the neural system destabilizes and responds with construction of new connections in order to better reflect the structure of the environment. Technologies, constructed niches (and prisms) have a causal role in this destabilization and modification. I'll move now to some more examples.

## Illustrative Examples of Constructive Development

Let's work through a set of examples of constructive processes in the brain. We have already seen the plasticity and potential for neural networks to grow (in barn owls), next I'll show that neural resources can be commandeered for completely novel purposes (in bird chicks). We'll see how altered neural inputs alter neural development in premature infants and in hand surgery. We'll look at the idea that cognitive devices can be transmitted culturally when circumstances are right, also at contextual impacts on mouse and rat brains. Finally, I will describe manipulations of human self-representation.

Lorenz (1937) performed imprinting experiments on bird chicks. Ducks and quails will

approach a speaker playing the maternal call of their species. At first it was thought that this behaviour was hard-wired (or 'innate'). However, it became clear that the behaviour results from prenatal exposure to species-typical calls heard through the shell wall (this is demonstrated by Gottleib's (1971) deprivation experiments). Interestingly, Lickliter (1990) then cut holes in shells three days before chicks hatched and shone patterns of light inside. The birds, who do not normally experience any visual stimulation except for diffuse light before birth, were stimulated with visual experience. These chicks did not approach any particular species calls once hatched. The prenatal *visual* experience had disrupted their normal *auditory* development. It was also shown to interfere with later auditory learning as well.

This begs immediately questions about human premature infant development. This is because very premature infants are thrust into a stimulating environment that is atypical for the stage of their brain development. Indeed, some studies have shown that the psychological development of premature babies is highly abnormal (Stephens & Vohr 2009, Johnson et al. 2009). Prenatal species-atypical patterns of stimulation (inducing developmental plasticity) have been implicated in elements of both mental retardation and genius in premature human children. This is not surprising given that the prenatal stimulation of Lickliter's chicks advanced their visual development at the expense of normal auditory development. Just as the juvenile owls co-opted uncommitted neural tissue to cope with the prisms, Lickliter's bird chicks co-opted 'auditory cortex' to cope with the visual patterns stimulating them in the shell. Constructive processes begin before birth. However, it is still important to note that parental education is still a better correlate of high IQ than birth weight or maturity. This suggests strong environmental influences on the development of intelligence (Weisglas-Kuperus et al. 2009).

Wilson (2010) notes that there is significant cultural variability in cognition. Culture influences the contents of cognition, but culture also influences the structure of cognition. Many 'cognitive tools' are not genetically wired, but are transmitted culturally and become 'firm-ware'. For example, number concepts are effectively transmitted via the inheritance of number words, and the algorithm for doing long-

division is passed on in similar fashion. Wilson discusses 're-tooling', but this is probably misleading as it implies there is something there (innate?) that is already tooled. Really she is talking about *organization*, another term for this is *development*. 'The consequence of this is that cognitive phenotypes will differ across cultures, sub-cultures, and even more local groupings such as families.' (Wilson, 2010, p. 1).

Wilson argues that we have a set of cognitive tools which underpin cognition. She argues, however, that many of these tools (the structure, not just the content of cognition) are culturally caused. Just like the prisms cause novel cognitive tools in juvenile owls, culture causes novel tools in humans. She argues: First, that cultural re-tooling of cognition is ubiquitous, and second, that cultural tools construct neural-cognitive architecture.

Wilson provides examples of this, such as the cognition of individuals who have 'unusual experience with spatial processing' (2010, p.4). London taxi drivers must complete years of memorization of the streets of London and this coupled with on-the-job experience give them fantastic abilities to navigate the city. Maguire et al. (2000, 2006) have demonstrated changes in the grey matter in the hippocampus that cannot be explained without reference to their navigation experiences. These taxi drivers also show a striking decrease in their ability to acquire new spatial information. Again it is interesting to compare this finding to Lickliter's bird chicks and the barn owls. Probably what has occurred is a degree of neural commitment. This is exactly what is predicted by the neuroconstructivist picture of cognitive development. We *often* see context-sensitive construction followed by a decrease in flexibility. However, a decrease in flexibility is not a *necessary* consequence of context-sensitive construction. Learning to read, for example, leads to an increase in plasticity in many domains.

Surgery provides another, fairly controllable, example of contextual modification. The somatosensory region of the brain acts as a map of the body surface. Anatomical structures such as limbs, hands and digits are represented in proportion to the number of nerves they have supplying them. The location of these maps in the cortex

is fairly predictable, but accumulating clinical observations have indicated that the brain is more plastic than previously thought. If a nerve is cut so that the input to the brain is lost then we see neurophysiological reorganization in this somatosensory cortex. Further, if sensory experiences change, rather than cease altogether, the same phenomena occurs. Surgical procedures on hands are always accompanied by synaptic reorganization in the somatosensory cortex (Lundborg 2000). Similarly, extensive use of hands (such as violin playing, and probably more recent phenomena like texting or typing) result in enlargement of the corresponding projectional areas in the brain (Schwenkreis et al. 2007). Many results of surgery, both encouraging and disappointing may be explained in this way. This process of reorganization is a system-wide phenomenon. Brains and brain regions exist in a context, and when that context changes, and the corresponding inputs to the brain or brain regions change, then structure changes.

Laboratory animals, mostly mice, raised in more complex environments exhibit behavioural, anatomical and molecular changes compared to mice housed in bare cages (Lewis 2004, Turner et al. 2003). The complex environments tend to consist of more social interactions, increased object density in spaces, and the provision of novelty. The outcomes include: increased spatial and non-spatial memory acquisition, decreases in voluntary alcohol consumption, decreases in aggressive behaviour, limitation of deleterious effects of maternal separation, and decreased stress reactivity. Increases in brain weight, cortical thickness and neural connection density were noted, as were increases in neural growth in the visual cortex and hippocampus and other molecular changes. In one of the best known ‘enriched mice’ studies (Cooper & Zubek 1958), rats were bred to be ‘maze-bright’ or ‘maze-dull’. Enriched environments caused maze-dull rats to perform better, but had little effect on maze-bright rats. Impoverished environments caused poor performances in maze-bright rats, but little effect on maze-dull. So these constructive interactions are very complex.

If we trim off one side of a rat’s whiskers, then within one to three days cortical responses to stimuli from the intact whiskers are strengthened (hence the brain’s

plasticity mechanisms are quick). So we see that altering inputs and outputs can dramatically alter cortical organization. The existence of the body, of a particular body organization, constrains the kind of experiences that are possible. Tool use can have similar effects on self-representation.

Operators of robotic devices handling hazardous materials at a distance come to 'rapidly and effectively' feel a shift in point of view, as if the robot arms are actually their own (Clark, 2003). The neural correlates underpinning self-representation have been shown to be malleable. When monkeys have been given tools such as a rake to reach for food, the cells in the visual receptive fields that represent the body become elongated along the axis of the tool after practice. Effectively, the tool becomes a part of the neural self for functional purposes (Iriki et al. 1996).

In an experiment conducted with humans Cardinali et al. (2009) demonstrate that after training with a tool for fifteen minutes, subjects perform poorly on the same task using just their hand. But interestingly, after tool use, the subjects localized touches on their middle finger and elbow as if they were further apart than they were. This suggests that the effects are partly caused by an updated body schema ('morphological updating') incorporating a longer arm. Given the monkey data just described, it is likely that visual receptive field neural modifications underpin this. Also, given the barn owls and prisms experiments, one wonders what would have happened had the subjects using tools been children, what if the experiements were conducted for seven weeks as in the barn owl case?

Mental representations of our own bodies are flexible. Ramachandran (1999) in particular has demonstrated this, often with innovative clinical applications. For example, stroke patients with paralysis can be coaxed back to voluntary movements through the use of mirrors set so that the patient thinks their paralysed hand is moving. Also, many illusions are possible. In one example, the subject's nose can be made to feel as if it is a meter long by stroking the nose of someone sitting in front of them, whilst their nose is simultaneously stimulated.

In the last fifteen years, experience with virtual avatars[6] has begun shaping the way that some humans experience themselves. The 'Proteus effect' is described by Yee & Bailenson (2009). In these experiments, people made to interact online using taller avatars negotiated more fiercely in online economic games than those using shorter avatars. Those with more attractive avatars behaved more pleasantly. Furthermore, and most importantly, these effects carried over to real-world interactions after the online games had finished. Other studies (Vasalou et al. 2007) show that private self-awareness is enhanced when avatars are similar to oneself. Again it would be very interesting to compare the effects of experiments like these with adult versus child or adolescent participants. I would expect more dramatic shifts of behaviour with children as was found with the adolescent owls.

It's not just that the blind man's stick or the sports star's racket, or the virtual surgical droid's grasping arm feel like they are a part of us, they literally become a part of us according to our neural representations. Clark (2003, p. 62) notes that, 'human brains seem to support highly negotiable body images'. Ramachandran (1999) likewise says, 'your body image, despite all its appearance of durability, is an entirely transitory internal construction'. Simple experiments can demonstrate this such as using a dummy hand to trick you into thinking you have sensations you do not have.

Just as the existence of a changing body (embodiment) constrains the experiences that the child can have at different stages of development, so too the technological world constrains experience. Children are active and different activity results in different experiences and different development.

Neural devices exist among a web of other neural devices. This is embrainment. If inputs to one brain device change, then the context of other devices changes and therefore the development of the child globally can change. It is essential to take constraints (and affordances) provided by brain, body and environment seriously when trying to explain cognitive development. Finally, more complex representations

[6] An avatar in this sense is a virtual, digital, movable, representation of oneself. The derivation is from the Hindu word meaning 'the worldly incarnation of a god'.

can be learned in higher-level networks based on simpler representations already learned lower in the network hierarchy. Societies gradually expose infants to problems they need to master. The result is that network-A confronts problem A, then it grows and network-AB confronts problem B. What is emerging here is a picture of representational construction, over time, that is flexible and employs additional cognitive resources as and when needed. However, we are also getting a picture of increasing specialization and commitment of neural tissue, which makes ongoing representational flexibility difficult. That said, the constructivist models show that complex representations (such as the owl's prism/ordinary vision coping device) can emerge without being pre-specified.

What seems to be going on here is that experience and context can alter the way the brain represents the agent, and a range of experiences can drive development of the brain in different directions. There are three ways that the brain can adapt during development.

1. The brain can construct new cognitive tools. We are seeing this with the examples so far, such as the barn owls and Lickliter's manipulations of bird chicks.
2. Second, the brain could find new ways to integrate existing neural devices.
3. Finally, the brain could integrate in novel ways with external resources.

In all three cases synaptic change is inevitable. Sometimes this may be just at the periphery of a neural circuit as existing circuits shift their pattern of integration, other times neural changes can be more fundamental and new or significantly re-wired circuits emerge. I've explained examples of (1) already. I'll look more at (3) in chapter three. For now I want to move to examining (2), the novel integration of existing neural devices.

## Cognitive re-tooling

Another way that cognition can develop in constructivist fashion is if existing devices are integrated in innovative ways according to what the context demands. It is

trivially true that we possess a suite of cognitive devices which underpin our thought and behaviour. Both Barrett (2009) and Dehaene and Cohen (2007) give accounts of cognition that explain the orchestration of complex cognitive accomplishments by appeal to this suite of more primitive cognitive devices. This is what neuro-constructivism predicts. Neural nets are built according to experience and then these devices can be integrated in many and novel ways to enable a wealth of cognitive tasks to be accomplished. Flexibility remains after neural resources have been commited to particular devices, but what is possible is constrained by earlier developmental outcomes.

Barrett (2009) argues that folk psychological categories map to neuronal firing patterns via an intermediate set of 'psychological primitives'. For example, various patterns of neural firing will contribute to a state such as core-affect, which will reflect a level of arousal or mood. These patterns need only be similar tokens of a type of state and not exactly the same in any given instance. Other types of states may be such things as directed attention, or categorization. Then a suite of such primitives can combine 'like a recipe' to produce myriad complex psychological phenomena the likes of which we call emotion, cognition and so on. This is a version of the type-token identity theory where folk psychological states, say 'anger', map to an array of primitives in combination, such as core affect, directed attention, categorization, and each of these maps to token brain states, which may be different in any given instance. The exact states will vary from person to person and situation to situation. Given the temporal nature of cognition, the metaphor probably ought to be a symphony of psychological primitives working in concert to produce cognition. Once the devices (neural tools) are constructed we can then construct ways of combining them to accomplish useful tasks. In other words, once we develop neural networks for sustaining primitive states such as affect, attention and so forth, we can combine these tools in different ways to construct novel cognitive states according to context.

Dehaene and Cohen (2007) have a similar argument. They call theirs the 'neuronal recycling hypothesis'. This is based on the idea that there exist a set of basic neural

networks these may be partly the result of evolutionary selection, and partly the result of developmental experience. The argument has three postulates, first, that human brain organization is subject to strong anatomical, connectional and evolutionary constraints. These constraints bias learning. Cultural influences on cognition, or 'acquisitions' (e.g. reading) must find a set of circuits that will support the required function, and are plastic enough to meld to the required form, and thirdly, that prior organization is never entirely erased. Obviously this theory is open to fine tuning of degrees. What is the degree of plasticity, the degree of prior organization, the degree of particular constraints? Basically, existing neural nets are recruited (in concert) to perform novel tasks as triggered by cultural experience. I take this to be a similar position to Wilson's, where the development of a suite of cognitive devices then permits novel cognitive processes to occur as these devices harmonize in various ways. Candidates for such primitives include the first two systems sustaining numerical cognition that were introduced in chapter one. Dehaene and Cohen (2007) take this further and note that reading, writing and arithmetic map onto remarkably invariant brain structures. These 'cultural maps' arise from transformation of cortical precursor maps present in other non-human primates. Culture influences brain development by exapting (through neuroconstructivist mechanisms) pre-existing brain systems.

Evidence from brain imaging studies suggests that as children acquire the ability to recognize words and perform algebraic calculations, that neural activation moves from a more diffuse pattern, involving the prefrontal cortex, to a more localized and specific pattern. Changing activation patterns seen with neural imaging suggest construction of cognitive devices rather than modular prespecification. As circuits are constructed they become more localized, specialized and efficient.

For example, studies suggest that when recognizing words a particular area of the brain, the visual word form area is reliably activated. This area responds more strongly to words in the reader's native script and to readable words and pseudowords than to random letter combinations. Interestingly, the activation response is larger and more bilateral in early readers and more specific and localized

in experienced readers.

The same phenomena is seen with numerical cognition where there is a progressive and massive transfer of activation from pre-frontal to posterior specialized circuits over time (Dehaene & Cohen 2007, also Qin et al. 2004). With mathematics the localizations are not as precise as with reading. This is to be expected because the inputs in reading are very precise, through particular transducers. The inputs constructing mathematical cognition are much more varied. They are cross-modal for one, we can hear numbers, and feel quantities as well as seeing them. However, the evidence suggests that our understanding of the cultural symbols of numbers is grounded (neurally) in links with neurons coding for specific non-symbolic numerosities in the intraparietal cortex (these are the primitive numerical systems discussed in chapter one, which are common to all primates). Furthermore:

> 'Several parallels between monkeys and humans suggest that the monkey intraparietal neural code for numerosity may be the evolutionary precursor onto which the human invention of arithmetic encroached' (Dehaene & Cohen 2007, p. 391).

Of course Dehaene & Cohen mean that the precursor was whatever system our last common ancestor with monkeys possessed. This is evidence, however, that for whatever reason, we possess some fundamental building blocks of cognition (such as number sense) that can be orchestrated in various ways to perform more complex cognitive tasks. This is the basis of the re-tooling, re-engineering, and cognitive integrationist accounts of cognition. But there is nothing to necessitate that these components themselves are innate, rather than the products of constrained development.

I read Dehaene and Cohen as having nativist leanings . One of their implicit assumptions is that our evolutionary history drives many of the constraints that exist in brains. For example they say that:

> 'the systematic difficulty that children exhibit in discriminating mirror-image letters such as p and q may ultimately be traced back to the native propensity of our visual

> object recognition system for mirror-image generalization, due to its *evolution* in a world where the left-right distinction is largely irrelevant' (2007, p. 385, my emphasis).

One wonders why *development* in such a world isn't enough to cause this confusion. Early development is indifferent to mirror-image objects. Developing brains would probably treat mirror image representations as equals unless their salience was noted by other concurrent causes and effects. The salience of written forms, and hence b,d and p,q distinctions, comes later, and requires updating of expectations or a re-modeling of constructed circuits. We shouldn't be too quick to apportion the causal power driving trait development to evolutionary selection for those psychological traits.

In fact some accounts of cortical function make very minimal appeal to evolution. Genes have a role in the general set-up of most organs, but these organs are usually fairly homogenous in structure. Kidneys have a repeating array of filtering nephron units that all function in much the same way. Lungs have a repeating array of alveolar sacs that all exchange gases in much the same way, and brains have a repeating array of cortical layers folds and columns. It seems likely that this, too, all functions in much the same way. And given the flexibility of constructive processes, this is likely to be some sort of general purpose learning device, which is open to experience. Not a richly pre-specified set of domain-specific tools. I suggest that we are the bearers of a cortex that is largely plastic, yet is so constrained by the body, the sensory apparatus, and the transducers converting sense data to neural data, that we end up with particular cortical architectures like the somatosensory map of the body's surface. If this is the case then both the nativists and the constructivists are right on one thing. We do have a common neural 'map'. However, this is because constructive processes are suitably constrained. The neocortex is organised comparatively late in development, and so has more degrees of freedom. Any invariantness is due to our roughly similar body-types, peripheral nervous systems, sub-cortical structures and roughly similar external developmental contexts. Any difference is due to differences in these things rather than pre-specified differences in neo-cortex.

Compared to other primates we are born very prematurely. In part this allows us to add experience to our endowment (Gilbert 2002, in Stotz 2010). In fact, it may be more accurate to say that we *substitute* experience for endowment, or we *overwrite* endowment with experience. It is likely that there are profound tradeoffs between what the brain was going to do, in a maturational sense, had we stayed in the womb for longer than nine months, and what it does instead after our encounter with rich experience.

Whether a cognitive device emerges due to adaptationistic genetic pre-specification, or through constructive processes, it is conceivable that such devices work in synchrony to accomplish additional cognitive tasks. Most nativists accept that environmental (and hence cultural) factors play some role in modulating cognitive phenotype. This view may be packaged as 're-tooling' or 're-engineering' the existing, species-typical, raw materials of the brain. Dehaene and Cohen's neuronal recycling hypothesis suggests that there exist a set of innate primitives, which are available to integrate in functional ways. I agree that there is a set of primitives, but neuroconstructivist phenomena suggest that these are likely to be unique constructions according to each individual's experience. I have hinted at one unjustified assumption that these author's have made in support of their nativist leanings, but now for completeness I'll give a summary of Evolutionary Psychology, which is representative of nativism, and emphasize two key criticisms that undermine it. Having done this, it should be clear that evolution, though probably constraining to some degree what can develop, does not determine our cognitive devices in most, or perhaps even in many cases. This will make sense of the neuroconstructivist examples I have outlined above, and the evidence of technological impact on psychological traits that I explained in chapter one. So let's look now at Evolutionary Psychology, the fallacy of adaptationism and the grain problem.

## Evolutionary Psychology

Evolutionary Psychology is a research program that emphasizes the evolved nature of the human mind and a universal human psychology (Barkow, Cosmides & Tooby

1992; Buss 1995; Pinker 1997). It should now be obvious why a universal human psychology is not a necessary consequence of an evolved human mind. The basic idea of Evolutionary Psychology is that the human mind/brain is a massively modular array of solutions to adaptive problems faced in the Pleistocene. Cumulative evolution by natural selection has driven information coding for these adaptations into the genome. The genome codes for psychological traits, though these may require developmental experience to emerge in species typical form.

An adaptive beahviour is one that is of use to the agent. An adaptation is a trait that was adaptive at some point and has been selected for over evolutionary time. It need not remain adaptive in the present environment. Evolutionary Psychology wants to avoid the mistake of positing current adaptive behaviours as adaptations, so it assumes that human beings have minds adapted to past environments. For humans the environment of evolutionary adaptedness was the Pleistocene. So humans have a 'stone-age mind' and consequently there is sometimes a mismatch between cognitive mechanisms and current environments. For example, it may be an evolutionary phenomenon that leads us to desire fat and salt rich foods. But when these are abundant (as today in the first world) this is maladaptive as it leads to illness. By modelling past environments, positing adaptive cognitive devices and testing current populations for those devices, Evolutionary Psychology claims to be able to work out specific cognitive systems and preferences shared across populations.

The evidence I have advanced above should make it clear that there will be no such thing as a 'universal human psychology' given the diversity of developmental contexts. But furthermore, we have seen that developmental environments that fall far outside what is species-typical or 'expected' by developmental resources can lead to dramatically different cognitive traits. Even if there actually were some evolved modular adaptations, then there is little reason to think that these would develop as expected in contemporary technological environments, given that these fall far outside what would have been species-typical during the Pleistocene. I will now outline two problems for anyone wanting to take an Evolutionary Psychology approach to explanation. These are the *fallacy of adaptationism* and the *grain*

*problem.* And then I will argue that even if there are some modular adaptations, that we could still see significant contextual variation in psychological traits as a result of engagement with technology.

### Adaptationism

One significant fallacy that evolutionary psychologists fall foul of is the following. Beginning with the premise that the mind can be studied naturalistically, and then taking methodological naturalism as the paradigm of investigation, they fuse it with our best theory of explaining the natural world, which is evolutionary biology. Up to this point I agree wholeheartedly with the move. However, evolutionary psychologists then seek to explain every cognitive trait as if it were an evolutionary adaptation. This is a mistake. The brain could be evolved and yet no psychological trait is an adaptation. The brain could be a general-purpose learning device. This adaptation *for learning* then underlies all other psychological traits, which are merely learned in particular contexts and are not adaptations (although they may be adaptive). Or one could argue that only some traits are adaptations and the many others are biological accidents of one sort or another. At the very least we will need to know what the brain is and how it works in order to accept or reject the hypotheses of Evolutionary Psychology.

Davies (2009) is representative of the savage attacks on Evolutionary Psychology in its most robust form. He charges the research program with failing to meet even minimum standards of scientific evidence. To begin with, life did not evolve in the way that Evolutionary Psychology claims it did. Many, if not most, biological features are not an adaptation argues Davies. There exist competing non-adaptive evolutionary hypotheses. Many psychological devices may be spandrels (Gould & Lewontin 1979), change resistant entrenchments, the results of drift, and so on. As Davies notes, the theses of Evolutionary Psychology are simply too strong, that all or nearly all of our psychological traits were *directly selected* for the performance of *specific* tasks. This in itself does not mean that Evolutionary Psychology is has nothing to offer, but it diminishes its explanatory scope. Nativist programs like Evolutionary

Psychology are simply not going to be able to account for all, or even most, of our psychological devices.

### The Grain Problem

There is a serious grain problem with the hypotheses of Evolutionary Psychology. The hypotheses of Evolutionary Psychology are lacking in reasonable evidential detail, or when there is reasonable evidence the hypotheses are necessarily vague as to be uninformative. Think of driving a car or reading a billboard. Many of the processes involved suggest the existence of a driving or reading module, a module being some version of a domain specific, mandatory, opaque, informationally encapsulated device (Fodor 1983, Carruthers 2006). Reading could conceivably be described as modular. Activity is automatic, we can't see glyphs on a sign as anything other than words, the knowledge is often encapsulated, and so on. But reading clearly does not begin as a module, a reading module, as such there is, develops with experience. So the existence of modules does not verify the Evolutionary Psychology picture, no one thinks there are driving or reading modules, which evolved in the Pleistocene.

Traits like driving competence are clearly built by experience in a novel environment. The adaptations, if indeed there are any, are of a different kind, such as the ability to alternate attentional resources, edge or depth perception, and so on. These are more general functions, more like the psychological primitives I have discussed under cognitive re-tooling above, which may be combined in novel ways to produce evolutionarily novel functions such as reading, or driving. Furthermore, it seems like such things as the ability to alternate attentional resources itself is built by experience in novel environments (see the computer games data in chapter one). Indeed, given the rapidly changing nature of the environment that humans have found themselves in over time, and using Evolutionary Psychology's methodology, it seems likely that the key selected adaptation would be the ability to rapidly learn new things, and not be bound by ancient constraints (there will be exceptions to this rule of course, as things like language, and cooperative behaviour will probably be adaptive in all imaginable human environments). This is the sort of picture favoured by Sterelny &

Griffiths (1999) and Sterelny in many subsequent works (e.g. 2003, 2011). I argue that this plasticity is exactly the core feature of the brain and human cognition. Much psychology may be transient, it appears in one generation persists for a while and then changes as cultural technologies and practices change and the developmental environment changes.

The grain problem continues when claiming that if a module is 'intact' then that subject has 'species-typical' cognition. This is simply a false premise as illustrated nicely by Karmiloff-Smith (2009) who examines Williams Syndrome, a developmental disorder that manifests largely as impairments in reading. Reading acquisition is disrupted, while apparently leaving other functions like face-recognition intact. However, although functional tests of face-recognition seem to confirm this, neuro-imaging seems to suggest that face-recognition in subjects with Williams Syndrome is highly atypical. At one level we have an 'intact' process but this tells us nothing about the highly varied instantiation in a particular individual. Standard tests often do not show how a process is being performed or if it is impaired in an individual given, say, an above average family base-line.

As I indicated in chapter one, Marr (1982) gives a very nice explanation of the different levels on which a process is operating. Some process like performing addition, making a pie, going shopping, or recognizing faces could be instantiated in many ways. There are three levels to understand:

1. 'What' the device is doing and 'why' (e.g. when calculating arithmetic, this is a mapping of two numbers to one). 'What' and 'why' *don't* depend on the specific system employed (e.g. Arabic numerals, Roman numbers, etc). The process itself puts constraints on the 'why'. For example when shopping if you buy two things it doesn't matter what order you present them, and if you buy nothing it should add to zero.
2. The second level involves choosing representations for the entities that the process manipulates. First, a representation for the input and output and second, an algorithm by which the process is achieved. This second level

specifies 'how' and *is* dependent on the specific system employed (e.g. Arabic numerals, Roman numbers, etc).

3. The device in which the process is to be realized physically. Obviously the same algorithm can be implemented in different technologies (e.g. an abacus, a calculator, a brain).

The choice of an algorithm is constrained by what it has to do and the hardware upon which it must run. For example, the number of columns an abacus possesses determines the magnitude of numbers that can be manipulated on it. Different algorithms tend to fail in radically different ways as they are pushed to their limits or deprived of crucial information. The same is probably true of face-recognition in Williams Syndrome, even though it appears 'intact'.

So 'intact' performance does not mean 'typical of species cognition'. There is impairment in development with subsequent compensation. Developmental context and constraints in Williams Syndrome have led to a functional 'module' with bizarre organization.

> 'Human intelligence is not a state (i.e. not a collection of static, built-in modules handed down by evolution and that can be intact or impaired. Rather, human intelligence is a process (i.e. the emergent property of dynamic multidirectional interactions between genes, brain, cognition, behaviour, and environment).' (Karmiloff-Smith, 2009, p. 61).

This is the neuroconstructivist picture and tends to suggest that there are merely different developmental trajectories. Even if there are modules that are adaptations this does not mean that they are developing as they previously have given contemporary contexts. And even if testing suggests an intact module, the neural organization, and perhaps the computational algorithm, may vary across populations. Even if Evolutionary Psychology finds a trait that looks genuinely universal, we cannot say that the function is implemented or developed in the same way across individuals. There may just be developmental convergence rather than an adaptation at work. A degree of developmental convergence across populations is likely because all children experience many of the same aspects of environment. This includes light,

sound, gravity, exposure to language and social interaction (Stiles 2009). Modules, as such there are, can develop rather than evolve. This indeed seems to be the case. Human universals are fully compatible with both nativist and constructivist accounts. But constructivist accounts do a better job of explaining observed variations in traits (Levy 2004).

It may be a mistake to absorb the full weight of arguments such as Davies' and Karmiloff-Smith's, or to deny modules outright, but now we see that there are at least five ways in which cultural technologies could conceivably influence cognitive development even within a strongly nativist picture.

1. Novel environmental inputs to a fixed native system could produce novel, or sometimes 'misfiring' functional outputs (this may be the case with pornography and sexual arousal, desires for salt and fat).
2. Existing coarse modules could be re-coordinated to produce novel traits like driving and reading that have no evolved history.
3. Developmental contexts may disrupt the development of modules, however, their functions may arise regardless through new developmental trajectories.
4. Modules that may once have existed may simply not develop in a novel developmental context.
5. Changing integration of modules with the external world of tools and symbols could cause novel functional profiles (see chapter three).

The point of all this is that, even if one accepted the premises of Evolutionary Psychology's arguments, there are still a great many ways in which technological context could drive psychological development. However, I suggest that the positive case for constructivism and the two objections to Evolutionary Psychology cast doubt on whether we should accept these premises.

The most recent concerted defence of Evolutionary Psychology is illuminating. Confer et al. (2010) concede that all evolved mechanisms require some environmental input for their activation.

> 'It does not make sense to ask whether calluses or mating decisions are "evolved" or "learned" or due to "nature" or "nurture". All evolved mechanisms require some environmental input for their activation' (p. 116).

They note that the key explanatory challenge is to identify the learning adaptations that enable humans to change behaviour in useful ways given certain environments. But the adaptations must be robust in the face of this developmental novelty or something else will develop. In this recent defence Evolutionary Psychology is taking a step toward more constructivist approaches to cognition. Approaches that take developmental context seriously and are not committed to either a universal or deeply 'innate' psychology. For the purposes of my argument I am happy to concede the existence of *some* evolved modular adaptations. Perhaps there is a language module for example. Although I think these are unlikely, see for example Cowie (1998), and Tomasello (2008). Also, Sterelny (2011) argues forcefully against an innate moral sense over and above our emotional dispositions. But even if there are some evolved modules, it must be agreed that developmental experience plays a large role in shaping the final nature of such devices. The brain is built for change and adaptive plasticity. That is its biological function. That is the adaptation. Any universalist project is misguided, and the notion of innateness is confused.

## Innateness Again

Mameli and Bateson (2006) argue that there is no simple correspondence between the folk concept of innateness and a scientifically useful definition. A number of different scientifically useful concepts partially overlap with the folk concept of innateness, but it is not possible to choose in principle among such concepts. Mameli and Bateson work through twenty-six possible candidates for a scientific successor to the folk conception of 'innateness'. They argue that none is problem free. Furthermore, after removing all candidates that are 'defective' versions of other candidates, they put forward eight possibilities. 'Innate' may mean:

1. Reliably appearing in a particular stage of the life cycle.

2. Being such that environmental manipulations capable of producing an alternative trait are evolutionarily abnormal.
3. Not produced by a mechanism adapted to map different environmental conditions onto different phenotypes and, at the same time, not produced by the imapct of evolutionarily abnormal environmental factors.
4. Generatively entrenched.
5. Developmentally environmentally canalized.
6. Post-developmentally environmentally canalized.
7. Species typical.
8. Standard Darwinian adaptation.

Mameli & Bateson note that the overlap between the folk concept 'innate' and this list of criteria is only partial. There is no principled way to decide which concept(s) most map to the folk concept and, importantly, the eight proposals are non-equivalent, they each include and exclude different things as being innate. Rather than focusing on outdated determinist accounts of innateness, Mameli & Bateson argue that what we need to know is if these eight 'i-properties' cluster or not. Is there a set of traits that have most of the i-properties and another set that have very few of them. In such a case then the i-property-rich traits will be the innate ones. These authors are sceptical that such a clustering exists. Further complicating this issue is that fact that several of the eight i-properties come in degrees. Traits can be more or less generatively entrenched for example.

The folk concept often seems to under-specify the complexity of biological mechanisms. What we really need is more emphasis on the idea that it is development rather than inheritance that is the central concept in biology. All evidence points to the importance of the manner in which inherited and contextual elements are combined as giving rise to one phenotype rather than another. Neither genes nor environments prescribe outcomes. It is confused to talk of traits being innate or non-innate, and evidence suggests that cognition is highly plastic and contextually malleable. However, I don't think all is lost for the 'innate' concept even

in light of the flexibility of the brain, the neuroconstructivist picture, problems with nativist accounts of cognition, and confusion surrounding the term 'innate'. I think there is a related concept, which draws on an analysis of the causes of traits, and the features of the privileged causes of any given trait. In chapter seven, after I have discussed causation in complex developing systems, I will introduce the idea of traits whose privileged causes are internal, stable and heritable. This account of 'innateness' hinges on an analysis of causal and explanatory privilege, which I will discuss in chapters four through six.

## Some Conclusions

I have argued that unique combinations of psychological traits arise in different individuals thanks to the context in which they develop. Context can alter neural growth patterns, the integration of different cognitive units, and in the next chapter I will show that context can afford different ways for brains to integrate with external resources. This context partly consists of human technologies. Technology can shape neural development. In the next chapter I will show how technology can support cognitive functions, and also sometimes ought to be seen as a proper component of extended, hybrid information processing systems. All these processes come together to give valuable insights into Human Nature. But Human Nature should not be confused with innateness (Stotz 2010).

I have spent two chapters building a positive case for context-dependent construction of psychological devices. We've seen how human intelligence, affect, and self-representation can be constructed. We've also seen how these processes are underpinned by dynamic modifications at the level of neurons and neural networks. Often the crucial contextual variables are technological. Artifacts, mathematical symbols and ways of interacting with digital virtual worlds contribute to these constructive processes. This complex picture will become even more complex in chapter three when we look beyond the brain, to its integration with the world. Ultimately we will need some principled explanatory method to untangle this web of

constructive development in order that we can articulate coherent explanation of the emergence of traits. I will address this complexity in chapter four.

Furthermore, even if there is evolutionary specification for a particular sort of cognition (thanks to the selective effects of a particular sort of environment) then changing the environment (as cultural technologies such as language, symbolism, numbers, literacy, digital technologies, simulations, formal education and so on, have dramatically done) will change the developmental trajectory. Even if some evolutionary specification is in the form of conditionals such as 'if social cue X then decrease androgen production, if social cue Y then increase it' (see Confer et al. 2010 for similar examples) it is still obvious that any particular environmental state can fall outside of the scope of such conditionals.

There are different ways that novel cognitive phenotypes occur. At the level of neurons we see stimuli that cause particular developmental trajectories rather than others. Lickliter (1990) demonstrates this with bird chicks. At the level of cognitive devices we see re-organization of individual (already developed) devices into a range of different functional units. Dehaene & Wilson's approaches to literacy and mathematical cognition are examples of this. Finally, we might also see expansion of particular neural networks if the context demanded more specific or a greater range of processing in a particular task. The prismed owl's developmental responses demonstrate this.

Whatever the mechanism employed in any particular case, altered environments can cause alterations in cognition. The take home message is that the neural processes I have described in this chapter explain how the cultural technological causes I outlined in chapter one can have the effects that they do on psychological traits.

Psychological traits are built by the interaction of, (a) plastic neural resources, (b) constructive developmental processes, (c) re-tooling or reorganization of fundamental processing units and, (d) integration across brains, bodies and technologies. The evolutionary aspects of psychology become relevant in constraining

the exact nature of (a) and perhaps in providing some primitives for (c) to work with. Overall we have a flexible and reactive genome and neural system.

Three quotes from Karola Stotz perhaps best sum up to this point. First: (2010, p. 488):

> '[W]hat is most distinctive about humans is the reaction of extremely developmentally plastic brains to a total immersion and involvement into a well engineered, cumulatively constructed cognitive-developmental niche (Clark 2008; Sterelny 2003)'.

In this Stotz is absolutely right. In this chapter I have pointed to the dynamics of the developmental plasticity and in chapter one I have pointed toward the total immersion in a cognitive developmental niche. Second, Stotz (2010, p. 498):

> 'Human nature must inevitably be a product of its cognitive-developmental niche that includes a great deal of cultural and symbolic scaffolding.'

Stotz unfortunately gives virtually no evidence that it actually *is* a product of this. Hopefully to this point (and continued in the next chapter) I have shown some of this. Finally:

> 'It is possible to wholeheartedly endorse the idea that the mind is a product of evolution without accepting the claim that the mind is constrained to develop or to reason in certain, limited ways. The key to separating these two claims is to recognise that what individuals inherit from their ancestors is not a mind, but the ability to develop a mind. The fertilised egg contains neither a "language acquisition device" nor a knowledge of the basic tenets of folk psychology. These features come into existence as the mind grows' (Griffiths and Stotz 2000, p. 31).

The brain is an evolutionary buffer system. It protects the body from the world and enables adaptive integration of body and world. Buffer systems are striking in that the buffer itself can change dramatically while preserving the integrity of the system. Such a buffer system is designed to maximize phenotypic plasticity and therefore permit adaptive behaviour in novel environments. This evolvability has probably been selected for over evolutionary time and is an adaptation. Such an adaptation provides

the basis for rapid change of human minds over short time-spans, even, as we have seen, within the space of one generation.

# CHAPTER THREE – Mind: Scaffolded and Extended

To summarize up to this point, in chapter one I described the vast array of examples of cultural processes that shape the way we think. In chapter two I outlined the likely brain and developmental mechanisms that permit culture to sculpt cognition. I now want to shift focus, away from the developmental causes of brain structure, which we might call scaffold-*ing* of development, towards seeing how the brain relates to external objects and incorporates them into functional systems. Such functional systems may be scaffold-*ed* in real time by external props.

Let me be very clear about this distinction. A scaffold is a supporting framework. Sometimes this is a temporary framework as in the scaffold used against a building to allow it to be constructed. Sometimes the scaffolds are permanent, as with a vine that is scaffolded by a trellis. Cognition could be (a) scaffold-*ed*, when it depends causally on external symbols to complete some operation, or could require (b) scaffold-*ing* for its development, after which operation no longer causally depends on the scaffold. An abacus could provide both versions of scaffold. When adding large quantities an abacus can be employed to scaffold the operation in real time, scaffold*ed* cognition, but also skilled abacus users have demonstrated very specific cognitive enhancements, such as increased digit span memory to around fifteen digits (Lock & Peters 1996). This enhanced span ability requires the scaffold*ing* of the abacus for its development.

In this chapter I will outline the theory of niche construction. These are the processes whereby organisms modify their environments in ways that have functional, developmental, and evolutionary consequences. Some author's have linked a hypothesis of extended cognitive processes to this theory of niche construction (e.g. Clark 2008, Stotz 2010, Sterelny 2010). We will see how modifications to humans' environments can lead to scaffolding of cognitive development and scaffolded cognition. I will also explore the further claim that external resources constitute

proper parts of systems that perform cognition, and in some limiting cases could be properly characterised as cognitive.

Not all cases where some environmental factor interacts with the cognitive system should be considered a systemic whole. For example, observing and responding to a billboard advertisement is a process of using an external representation, but the representation is merely taken as an input to the cognitive system. However, it is argued that the causal coupling of inner and outer elements sometimes creates a 'sufficiently integrated cognitive whole' (Clark 2008, p. 74), these are the cases of extended mind, which I will examine presently. Sometimes understanding the cognitive system as actually extending beyond the body and brain helps us to explain surprising facts.

## A Brief Introduction to Niche Construction

Niche construction (see Odling-Smee, Feldman & Laland 2003; Sterelny 2003) is the process whereby organisms modify the environment. Environments may be modified in adaptive ways, such as a beaver building a dam, or in maladaptive ways, such as humans producing vast quantities of atmospheric carbon dioxide. Niche construction can have four important consequences. First, the altered niche can effect development. Second, the altered niche can contribute support to function. Third, the altered niche can constitute a novel evolutionary selective environment. Fourth, the altered niche can drive further change and/or accumulate over time. Putting aside the evolutionary selective consequences of niche construction for a moment (I will return to evolution at the end of this chapter, and in chapter eight) let's look at some examples of the other general consequences of niche construction.

A nice non-human example of the first consequence of niche construction, and a case of the scaffold*ing* of development, is exhibited by meerkats. They ensure their offspring can safely learn how to catch scorpions, first by providing dead scorpions to play with, then by providing disabled (stingless) scorpions, and then finally fully functional scorpions (Thornton & McAuliffe 2006). The parent meerkats are

engineering the developmental educational environment of their offspring. Sterelny (2011) gives many examples of scaffolding of the development of human hunter-gatherer skills based on an apprentice model. Humans engineer their environments with the result that human children learn stepwise complex behaviours and cognitive skills. Perhaps the most dramatic demonstration of this process is formal education where students are led through a series of demonstrative exercises, often with partially worked examples, and learn, piecemeal, complex cognitive manipulations. Skills are gained that would likely not have been gained had the developmental niche not been engineered in this way. Importantly, it is often the case that had the niche been engineered differently, different traits would have developed.

The second consequence is that changes to the niche can be functional. Think of doing long division. Without a suite of niche engineering this is impossible for most people. The process is supported by pens, pencils, or chalk, or some other marking tool. The process also requires paper, or a whiteboard, or some other surface to make marks on. Finally, an algorithm is needed. This may be internalized but that is not essential, a written algorithm to be followed in real time would suffice. Perhaps, as is the case more recently, the process is supported just by a calculator or computer. In any case, we have engineered our environment in such ways that we have as needed, when needed, access to all these tools and algorithms. The niche we have built supports this mathematical cognitive function.

Alterations to the niche can change or accumulate over time. Development doesn't stop. It is a lifelong process and so environmental engineering can scaffold lifelong development. Also, scaffolding doesn't have to be intentional. An example of this arises in the context of formal education in mathematical physics. Formal educational technologies are some of the most powerful cognitive transformers that we have invented. Wimsatt and Griesemer (2007), drawing on Warwick (2003), outline the development and effect that the Cambridge University tutorial system in physics has had on enhancing physical thought. In 1750 due to a desire for increased objectivity in the examination process Cambridge changed its examination system in mathematical physics from an oral system to a system employing written exams (a

change in the niche). This altered the nature of the material that was assessable from philosophical discussions of texts like Newton's *Principia*, to written problems assessing technical competence, for example, asking students to present proofs. This meant that overall ranking of the students became possible, so students were then under pressure to perform well. Many students began seeking external tutors to help them. The performance of the tutors was judged by how well their students did in the exams and so the tutors began to experiment with different techniques. One tutor made his class take notes (even though paper was very expensive) in order that they may practice writing fast for the examinations. However, as a side effect of this niche construction activity, these students, now armed with a permanent record of the teachings and problems, mastered the techniques much more easily than other students. In fact all of the following advanced teaching methods emerged at this time: interactive questioning rather than passive lecturing, guided problem solving, regular exams, textbooks, graded exercises giving feedback, and grouping students by ability so they could learn from each other. The university began to ask the tutors to set more challenging questions as they now had greater mastery of mathematical physics than the internal lecturers. Eventually, embarrassed by the competence of external freelance tutors Cambridge University internalized the tutorial system. Ultimately, a change from studying a set text philosophically, to practicing problems, led to solving problems never even contemplated by Newton. All this took place within the framework of a formal education system, which changed bit by bit over a period of 150 years. In math and science there is often a 'strong sequential dependence of later methods and results on earlier methods and results' (Wimsatt & Griesemer 2007, p. 232). It very much seems to be the case that as we build our worlds we build our minds.

## Embodied and Embedded Mind: Extending Functionalism

Minds (and therefore psychological traits) are embodied and embedded in an environmental context. 'Minds are not disembodied logical reasoning devices' (Clark 1997, p. 1). This simple truism has been overlooked in past philosophy of mind and cognitive science. We use our fingers to count with. We preserve culture, history and

memory in vast treatises for sharing information with others. Minds are inextricably bound up with bodies, cultures and environments. This has genuine implications for function.

Functionalism about minds is the claim that mental states are individuated by the causal relations they bear to sensory inputs, behavioural outputs, and other mental states (Putnam 1975). A mental state is what it is because it plays the right sort of functional role in the cognitive system. Traditional functionalism permits us to attribute minds to such things as aliens, super-intelligent robots and the like. This move is permitted because functionalism is not committed to the material nature of mental states and allows mental states to be multiply realized, in different materials, for example silicon, or in different arrangements, such as pulses of green goo. Some of these entities may be external objects.

External factors are necessary for cognition as we know it. In fact, external factors of the right sort might constitute parts of a functional system that thinks (Wheeler 2010). Wheeler argues that functionalism about minds is the correct approach and that we can have extended functionalism in some cases. This has the potential to give us 'extended minds' (Clark & Chalmers 1998, Clark 2008, see also Donald 1991).

Taking functionalism as our baseline for characterizing mental states, and accepting that there is no *a priori* reason to exclude external factors, then we see that some external props could constitute cognitive states rather than merely being objects that cognitive states are causally dependent on. Wheeler points out that there are two sorts of external prop.

> 'Bare causal dependence of mentality on external factors – even when that causal dependence is of the "necessary" kind just highlighted – is simply not enough for genuine cognitive extension. What is needed is the *constitutive* dependence of mentality on external factors... *for supporters of the merely embodied-embedded view of mind*, the pen and paper system is to be conceived as a noncognitive environmental prop'. (Wheeler, 2010, p. 246)

So, we motivate a debate over the bounds of cognition by asking, can we justify the transition from merely embodied and embedded approaches to an approach extending the mind? To accept that human cognitive systems are embodied and are embedded in a world of props, is not necessarily to accept that those props constitute parts of our cognitive systems rather than mere inputs or tools. To argue that external features are parts of a cognitive system is to argue for the thesis of extended cognition. Before we get into that debate, however, I would like to prime the reader with a case study detailing the origins of writing, since written symbols are often argued to be elements of extended cognitive processes (e.g. Clark & Chalmers 1998).

## Literacy

I will begin by outlining some of the details surrounding the emergence and effects of literacy in humans. Writing lies somewhere in between 'a system of human intercommunication by means of conventional visible marks' (Gelb 1963, p. 12) and, 'the graphic counterpart of speech (Diringer 1968, p. 8). Writing systems originated approximately 5000 years ago, and depended on pre-existing visual/iconic and auditory/symbolic cognitive systems. But necessary, too, were a range of technologies, tools, marking materials, surfaces for writing to persist on, and so on. The *motivation* to make durable marks was also necessary. This motivation emerged from a social problem space. With agriculture, societies had become spatially and temporally extended. There was also intensification of the organisation of the social environment. Biological human memory became challenged, property was accumulated and there was more to keep track of. Information had become more valuable and information in an oral culture was fragile (Lock & Gers 2011). Another way to put this is that the bandwidth and fidelity of oral culture was not up to the task demanded of keeping records in larger agricultural cultures.

Lock & Gers (2011) further argue that, just as is the case with the emergence of tutorials at Cambridge University, no one set out with the intention of writing speech down. Contemporary systems of coding speech evolved through a process of natural selection once visible marks were initiated to confront the problem domains

identified above. Just as marking tallys enhanced human capacities for remembering and tracking quantities, such as the lunar month and seasons, and pictographic and cunieform symbol systems expanded accounting and trade practices, writing down speech massively enhanced what it was possible for humans to do in the world. Writing down speech was not the adaptive problem that early notation systems set out to solve, but this adaptive space was opened to humans once this process began. Simple mathematical problems and tallying seem to have created a mind space for literacy. In turn literacy builds a host of other cognitive skills.

Vygotsky argued that language influences the higher psychological processes (1962). This is corroborated by Pica et al. (2004) and Everett (2005). If this is so then there seems to be 'an *a priori* case for assuming that subsequent changes in the means and modes of communication would affect cognitive processes in parallel ways' (Goody 1987, p. 260). A similar line has been taken by McLuhan (1962) who makes very strong claims for the cognitive consequences of literacy, particularly the alphabetic system of writing. These claims need to be approached carefully. For example, multiplication is much more easily conducted with the Arabic as opposed to the Roman system of notating numbers. But easier doesn't mean different: both systems enable the same underlying cognitive process – the manipulation of symbols for arithmetic purposes. This may be true for simple examples, but the Arabatic algorithms are much more open-ended and error-free. Similarly, the fact that it is much more difficult to organise a dictionary of Chinese characters than alphabetic words does not imply that there are different cognitive processes involved in devising an organisational framework. On the other hand, there do appear to be clear biases amongst occidental versus oriental cultures in the way they approach perceptual and cognitive problems (e.g., Nisbett, 2003).

We need to clarify what is involved here. To what degree are we interested in what building a literacy niche lets us do (scaffold-*ed*), and to what degree in how it transforms us (scaffold-*ing*)? There are three different positions one could take regarding the effects of writing on cognition.

1. There is a strong claim that the appearance of writing systems, and the associated memorial record that is open to public scrutiny, began to shape the cognition of those who learned and used them. Goody and Watt (1963) claim that there are changes in *kind* of cognition and critical examination. For example, Goody claims that the ability to write statements down allows the emergence of syllogistic reasoning. This encourages skepticism as a routine mode of thought. The idea is that certain cognitive logics depend upon writing. Olson (2005) suggests that this is because literacy allows metarepresentation of language. We are able to think *about* language not just with it.
2. Those who deny that writing has cognitive effects argue that symbols are merely tools that brains can use in order to perform tasks that they could do, in theory, internally. For example, Masterson (1970) describes twenty-one different uses of 'paradigm' in Kuhn's *The Structure of Scientific Revolutions*, and catalogues these. Such a task would be very difficult without literacy, indeed it may place impossible pressures upon the functions of biological memory, but the mental processes employed are of the same kind in each case (Halverson 1992).
3. A somewhat intermediate claim, which is that writing and written symbols form proper parts of *distributed cognitive systems* and *sometimes* play important constitutive roles in our thought processes.

We find defence of the cognitive differences claim (1) in Goody and Watt (1963). Goody recalls that he could not count cowrie shells as well as Ghanaian boys who are well practiced at this task, which is routine in their daily lives. But Goody notes that he could multiply the numbers of shells faster. The argument is that written times tables helped him learn and permitted rapid recall of visually inspected charts as needed. If the example is unconvincing, then a more illuminating claim may be that it is extremely difficult to form the concept of a negative quantity without having previously encountered a physical number line.

A further example provided by Goody (1977) is that of lists. Though lists exist in non-literate cultures, writing encourages the ability to inspect lists and isolate items for formal taxonomic or categorical purposes. One could argue that such categorical thinking is a new mode of thought. Donald (1991) describes human interaction with symbolic media as plugging into, and becoming part of, an external system. He argues that grouping of information into clusters, or lists, is a peculiarly visual institution. Visual lists differ from oral ones in that visual lists free up working memory. Once 'free,' working memory and attention can be directed to other tasks. This allows lists to be used without stressing the resources of working memory. Inspection and processing of the lists is possible without sustaining them within the resource limitations of working memory. Donald further notes that an alphabetically written word may be a phonogram (representing sounds), but it can also be an ideogram (representing an idea), but that even a whole paragraph or entire book can be an ideogram. We can then manipulate ideograms to produce new content.

> 'Each time the brain carries out an operation in concert with the external symbolic storage system, it becomes part of a network. Its memory structure is temporarily altered; and the locus of cognitive control changes' (1991, p. 312).

The reading of writing also requires a new suite of visual scanning techniques. Ordinarily we don't scan visual scenes from left to right, top to bottom (as English speakers do when reading). Furthermore, we can break information into chapters, parts, boxes, tables, sections, relations. We give ourselves new perceptual objects with which to interact in the world. Biological memory does not easily lend itself to this. Finally, Donald notes that, with writing, iteration and recursion are truly unlimited.

Numerically literate people perform calculations in a way that is qualitatively different from the multiplications of non-literates. Interestingly, the Arabic numeral system has evolved to be more adapted to this task than systems such as Roman numerals. This is not to say that one cannot perform complex calculations with Roman numerals (as the Romans themselves clearly did) but that there are differences in the algorithms employed. One way we perform mental calculations is

by internalizing these algorithms and this can lead to differences in the cognitive processing of number even when the symbols are not actually written down. This latter example is a case of scaffold*ing* of cognition, which results from repeated experiences of cognition that is scaffold*ed* by symbols.

Goody (1987, p. 256) further claims:

> 'the very nature of formal reasoning as we usually understand it (that is, in terms of Aristotelian 'logical' procedures) is a highly specific skill, critically dependent upon the existence of writing and a written tradition which helps formalize intellectual procedures'.

Logic is not impossible without literacy, but in general people struggle to develop logical thought without the ability to visually inspect. Indeed, Sperber (2000) argues that we have a long-standing, evolved, 'logical module' and that this allows us to attend to the logical properties of representations. This sort of device perhaps evolved because we needed to attend to the indefinitely many possible mental states of conspecifics and the inferences that they would likely make from them. It would also be essential within the context of a persuasion/counter-persuasion arms race of the sort likely to have emerged with language. Such a cognitive device would be content neutral, merely attending to logico-semantic relationships such as entailment and warrant. However, even Sperber concedes that this would be unlikely to account for the full range of our modern capabilities:

> 'The type of metapsychological competence that is likely to be an evolved adaptation is unlikely to explain the modern human ability to attend to abstract representations, and in particular to do formal logic' (Sperber 2000, p. 12).

This is likely to be true, given the demands on working memory of any moderately complex logical reasoning. It seems that written symbols are an essential component of serious logical cognition. In such cases as the analysis of 'paradigm' in Kuhn's writings, it may be true that the elementary operations are the same as those we have been capable of for a long time, but by relaxing the demands on memory and attention then new combinations and composites are possible. One could liken this to the transition from single-celled to multi-cellular organisms. The fundamental

building blocks are the same, but many new, varied and powerful functions become possible.

However, others are skeptical and instead support claim (2), symbols are merely tools. Halverson (1992) notes that a cumulative intellectual tradition is indeed aided immensely by writing, but that a syllogism is just a sequence of statements about relations that do not depend on the medium that the statements are presented in. 'Written records… may allow us to be more accurate in certain judgments, but skeptical attitudes hardly depend on them' (p. 309). He also argues that although lists abounded in ancient literate civilizations for administrative purposes, they do not seem to reveal any notable differences in kinds of cognition.

Halverson describes many phenomena that are only possible with literacy. These include, crossword puzzles, grammars, dictionaries, reading maps, and shaped verse. Literacy significantly opens the door to a whole suite of possibilities. Indeed, Nickerson (2005) concludes that, 'there is no other technological advance whose effects on human history rival those of this one' (p. 25). He catalogues a host of cognitive amplification devices ranging from various kinds of slide-rules to computers and memory aids, all of which amplify our capabilities.

It may not be possible or productive to distinguish between a major enhancement of existing capacities and the establishment of qualitatively new capacities. It is clear that literacy powers enhancements, illustrated by the analysis of Kuhn's text. Furthermore, when thinking, we often gesticulate and produce overt or covert linguistic utterances. Sometimes we write things down. These actions allow us to attend to particular aspects of thoughts, or to create new variety to stimulate further reflection. The thinking process is partly constituted by our internal representations and external scaffolds that we embed ourselves in. Clark often cites the example of physicist Richard Feynman. When told that his extensive notes and scribbling were the record of his work Feynman replied, 'No, it's not a record, not really. It's working. You have to work on paper and this is the paper, ok?' (Gleick 1992, p.409). Feynman is pointing out that his thoughts are hybrid, partly constituted by internal

representations and partly by his external working. An important difference between Feynman and prehistoric people is Feynman's integration with external media, mathematical symbols, pencil and paper, and graphic representations of concepts. Enhancement seems to shade into qualitatively new capacities.

A problem in presenting this as a debate between those who see literacy driving qualitative cognitive change and those who don't is that the debate focuses too sharply on individualistic cognitive processes. The Feynman example can be writ large. Literacy, through allowing thoughts to be 'frozen' permits new kinds of cognitive collaboration across space and time. We see new forms of the division of intellectual labour and new forms of teaching. A nice example of this is presented by Rambusch et al. (2004). In a study of artifacts in a hospital setting, these authors observe the role that the folder of patient notes plays in mediating between the cognitive activities of various staff, such as nurses, doctors, allied health professionals, and administrators. Each agent interacts with the written notes, reading off representations, engaging in activity, altering representations or elaborating on them. What emerges is a dynamic record of the patient's condition, treatment, management plan, lists of things to do or observe, a record of the thought processes of medical staff, and so on. Human cognition is a cultural and social process and work done by individual agents produces and alters this representational artifact. The folder accumulates information that no single person could remember and the current state of the artifact at any given time causes different individuals to perform different cognitive acts. Even the location of the artifact causes certain agents to perform certain acts depending on their cognitive role, as doctors, nurses, administrators and so on. We see processes of shared memory, of trust and coordination across a complex cognitive task, there is rich inter-agent communication as well, which hinges on the presence of a reliable, durable and modifiable store of represenations. Much of this would not be possible if the agents were not literate.

Someone might object that it is now difficult to individuate cognitive systems. If different agents are constantly interfacing with the tool, where does one cognitive system begin and another end? It is also difficult to determine which parts of the

system are cognitive, or are mental states, and which are not. However, this is not really important. What is important is to recognize that at any given time there is a looping process running from agent to artifact and back again. There is a process that involves both external and internal representations, and manipulation of those representations. These active processing loops are more than just input-output couplings, they are dynamic processes that are causally necessary for and support human cognitive action. These distributed systems enable human functions and enhance our capabilities in the world.

Even if there are no deep 'cognitive' (read internal, neural) differences between literates and non-literates, the emergence of writing certainly provided powerful strategies enhancing learning, memory and education. The ability to record and reflect upon content, to critically deconstruct long and complex texts, and to easily categorize the world, providing new objects for our attention in the form of lists, has amplified human cognitive prowess in important ways. The effect is that of a feedback loop, so that non-literates become progressively left behind.

The important idea is not that literacy causes important *large* changes in cognition, rather that small influences act through processes of accumulation and amplification. This led humans to think about things that they would not, without the emergence of writing systems, otherwise have thought about, or even been able to conceive. It is enhancements to the *ease* of formal logical thought, the scrutiny encouraged by written records, and the complexity of argument permitted, which are the important factors. A clear example is that early tallying does not require any abstract concept of number, but early tallying was probably necessary in allowing this concept to be achieved.

In the remainder of this chapter I will discuss Clark's notion of an extended mind. I argue that humans have hybrid thoughts, which are partly dependant on internal neural processes and partly constituted by external or potentially external things such as words and symbols. This introduces us to the idea that minds may not be entirely constituted by brains. We do indeed create surrogate contexts for our brains to

operate in, and these contexts sometimes constitute proper parts of cognitive systems. I will give examples. I then note that such contexts are cultural in nature and that even without shaping the development of brains, through niche construction, culture can shape the trajectory of the evolution of human minds. This may mean that the human mind becomes a rather vague and fluid entity of study, however, we can certainly still identify assemblies of physical things, which constitute various psychological functions. The study of these assemblies then constitutes cognitive science. Minds and psychological traits are not the same thing, though there is some relation between them. Perhaps minds are collections of psychological traits. The argument here is merely that psychological traits, i.e. what humans can do or think about, are not causally dependent exclusively on brains. The sustaining systems are hybrid in constitution.

## Hybrid Thoughts

I will now spend some time discussing the notion of a hybrid thought. When something is hybrid it is partly constituted by one kind of thing, and partly by another. A hybrid vehicle is powered partly by petroleum and partly by batteries. A hybrid university course is partly delivered as face-to-face classes and partly online. In these cases there are (at least) two different kinds of thing, which act in complimentary fashion to produce some outcome. Clark argues that the human mind is naturally designed so as to 'co-opt a mounting cascade of extra-neural elements as (quite literally) parts of extended and distributed cognitive processes.' (2006, p. 1). Clark calls this our 'ancient trick'. I hope it is clear from chapter two that this is exactly the sort of thing that the processes of neuroconstructivism are geared to do. Once this very generalist mechanism is in place, then brains find innovative ways to integrate an evolving cultural context with neural systems.

As I have outlined in chapter one, our numerical capabilities depend on three systems; a small quantity individuation system and an approximate magnitude system (both of which are biologically primitive). We also have the learnt capacity to use number words (this is an instance of re-tooling our primitive number systems).

This is not biologically basic. We learn that perceptually distinct number words correlate to distinct quantities, without having to represent the quantities themselves. We can then perform actions like working out long-multiplication on paper.

When calculating the product of two large numbers algorithmically, on paper, there is an embedded, dynamic, causal coupling between my brain and the symbols. The *process* of arriving at the final answer is constituted by my brain and external symbols. But, this is not mere embodiment or embedding of a cognitive process in an external environment of tools, because there are representations beyond the limits of my body, crucial to the process, that are not internalized in any way. We only multiply single digits. When multiplying 382 by 112, I begin by multiplying two times two. And yet multiple digit solutions resolve themselves on the page. The product of a long multiplication is not represented internally at any stage of the algorithmic process. Yet it becomes available to internal processes for representation at the end of the working. The final product is represented by the final row of digits in the multiplication and is a representation now available to further cognitive processes. I am manipulating single digit symbols all the way through, yet the manipulation of larger numbers emerges in this process. This is a hybrid process, partly relying on internal neural operations and partly relying on external scaffolds and support. The argument is that lots of mathematics is emergent and cannot be reduced to brain-bound, symbolic, or cognitive subsystems.

Once the manipulation of representational vehicles is seen as part of the process of thought, then it is hard to delineate external from internal without applying a neural chauvinism over what constitutes cognition (Menary 2007). Once people have writing systems available to them, cognition seems to bleed into the environment as integrated systems of information processing. This is qualitatively different from non-literate thought, it is hybrid.

One objection to this approach would be to cite instances of humans who can calculate such large multiplications in their heads. The savant Daniel Tammet is one

example, by visualizing numbers as coloured shapes which merge together he calculates products in seconds that would take ordinary people many minutes on paper. Other individuals can extract $13^{th}$ roots from one hundred digit numbers in their head. However, there are many ways to perform certain cognitive tasks. This is the whole basis of multiple realizability. We could employ a calculator, we could physically count out beans, or we could do the sums in our head. Each is a different instantiation of the arithmetic process. Some instances are examples of traditional internal cognition, some instances are genuinely hybrid. It does not matter for the arguments here if some people can perform a task using thoughts that are never hybrid, because it is clear from observing the behaviour of a vast majority of others that hybridity is a normal and important mode of mathematical reasoning. It is unclear how savants manage to perform the feats that they do, what is more clear is that ordinary human cognition is often hybrid and dependent on external structures.

Negative quantities are a good case study for hybrid and extended thoughts. For example, the Greeks based math on geometry and so had no negative numbers. Negative spaces, of course, cannot exist. The Chinese had a red and black rod system which represented who owes and who is owed money. It could be argued that this was a representation of negative quantity, as too it could be argued, were the Indian Brahmagupta's 'rules for zero' which employed concepts of fortune and debt. However, true negative values must be decoupled from sums of money and the social construction that is debt. Debt is still a positive value, a sum of money owed to someone. It wasn't until Europe between 1500-1900 that mathematicians began to use negative numbers for solving quadratic equations and where the roots of negative numbers had to be dealt with. The imaginary number *i* was invented. *i* cannot be translated into an internal representation in the same way as a discrete real number. The external symbol *i* becomes a placeholder, a focus of attention until it can be eliminated by manipulation. This is a truly hybrid thought. It could be argued that the parabola $y=x^2$ could be modified with a third 'z' axis running perpendicular to both x and y axes. This could enable a visual representation of an inverted 'ghost' parabola representing negative roots. However, with this it is still at least the case that the media is required for priming the internal representation of *i*. 'Since

mathematicians can express operations that can never be performed in the mind alone… mathematics can convey a range of ideas that are actually opaque,' (De Cruz 2008). These are true hybrid thoughts. The claim is that there is (at least) a stepping stone in development where a hybrid thought acts as necessary scaffolding for an internal concept to form. That hybrid thought is based on a public symbol.

De Cruz (2008), like Clark, is confident that there are hybrid thoughts. 'External symbolic representations of natural numbers are not merely converted into an inner code; they remain an important and *irreducible* part of our numerical cognition.' (De Cruz, 2008, pg 487). Public media certainly play an essential role ontogenetically and operationally, for example, by allowing a mathematical thought to proceed without having to represent all the content at once.

Many of us have experienced the formulation of an elusive thought as we have begun to write about a topic. Often an idea may not be clear until we begin to write it down, and only then reflect on the words, their logical form, and notice that the idea lacks clarity. This occurs because once we begin to produce public representations we can attend to particular concepts and manipulate hybrid thoughts on the page by engaging with the material symbols. This idea leads into a second, non-mathematical example of hybrid cognitive processes.

When I was writing this chapter, I performed the following procedure. I read a pile of articles and books on relevant topics. I then typed up notes and comments pertaining to each article or book. I also wrote some chunks of text that consolidated some of my thoughts. I ended up with a large word document of some ten or twelve thousand words. This was a document that I could not wield all at once, let alone fit on the screen. I struggled to work with it in this format. So I printed it out, and took a pair of scissors and physically cut all the bits up and arranged them by topic and ordered the topics into an argument. I then took each new section and wrote them up on my computer again. This was *hybrid* cognition. By physically manipulating the environment and especially manipulating objects that were 'frozen thoughts' which I could stop thinking about because they were frozen in media, I greatly simplified the

process. The entire system of my brain, body, scissors, paper, and symbols, all these *complimentary* entities, *constituted* the process of writing this chapter (see figure 5).

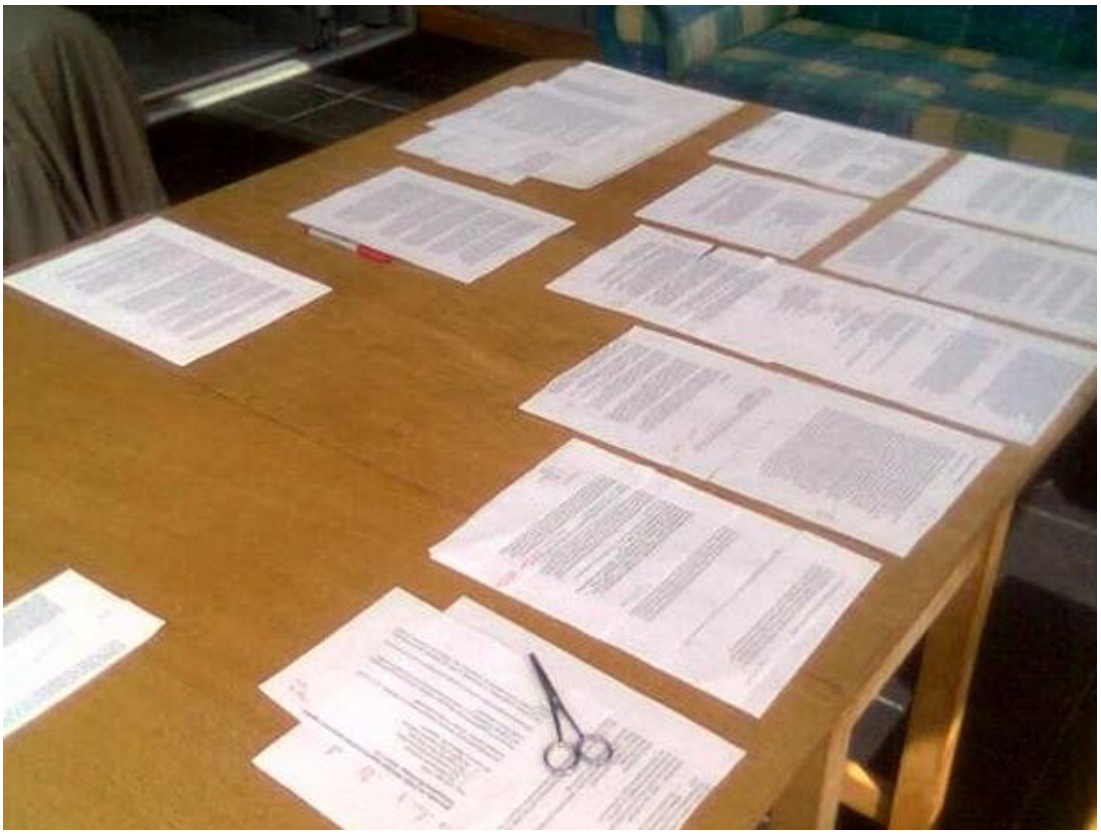


Figure 5: My hybrid mind while writing this chapter.

I am not arguing that it is impossible to compose such a chapter (or indeed solve mathematical problems) without hybrid thought processes. I am arguing that in actual fact, through hybridizing, the process is made much more simple, or is avaibable to individuals not capable of performing the task entirely in their mind, or that much more complex psychological functions can be instantiated.

One important point I would like to highlight here is that irrespective of the ontological hybridity of thoughts, there is an important causal dependence of many of these sophisticated thoughts on external media, the body, and our integration with the world. The notion of hybridity itself is less important than the causal structure of the systems that underpin psychological traits. That said, the notion of hybridity helps to focus our attention on important and central cultural technological elements of cognitive systems that may otherwise have been taken for granted.

Let's take just the example of language. There are at least three ways in which language (symbols, tags, physical media) can function in hybrid fashion:

1. Language as a source of additional targets for attention and learning (remember the tag trained chimps from chapter one). These may be cueing differences, by having two different symbols then agents can learn that there is a difference.
2. Language as a resource for directing and maintaining attention on complex conjoined cues, a label.
3. Language as providing some of the proper parts of hybrid thoughts (as we saw in the cases of numerical representation).

So by employing language I can anchor my thoughts. I can produce a linguistic symbol describing, for example, my mental state, 'I am sad'. I can then reflect upon this, my thoughts take shape thanks to a process of looping into the world and perceiving what I have produced. This is overt in the case of a Facebook webpage, for example, where I represent myself externally, perhaps linguistically, perhaps as an avatar in the three-dimensional virtual world of Second Life. These external props provide anchors for further reflection and contemplation, enhancing self-awareness (Vasalou, Joinson and Pitt, 2007). I outlined evidence of this sort of phenomenon in chapter one.

I have given examples of how public language, symbols, gestures and objects can turn thoughts into properly hybrid things. But we can take things further than mere hybrid thoughts. I will now examine Clark & Chalmers' (1998) stronger thesis that some cognition is essentially and significantly extended into the world outside our bodies.

## Extended Mind

Clark and Chalmers (1998) introduced the concept of an extended mind. This is an idea that in some forms goes back at least to Merlin Donald's *Origins of the Modern Mind* (1991). Extended Mind is also called vehicle externalism. Clark makes an

important distinction between the contents of thoughts and the vehicles of thoughts. The vehicle is the physical instantiation that permits the thought to exist, such as a burst of neural activity. For example, the perceptual content of 'greenness' detected at some location in visual space will have a pattern of neural firing as its vehicle. None of this neural firing is green. So there is a difference between the content and the vehicle. The content is the thought itself, e.g. an experience of greenness. Though we can localize the vehicle in physical space, we cannot really localize the thought. The content of a memory may be {the location of the Museum of Modern Art} but the vehicle for this might be some neuronal state in the brain, or the symbols in a notebook. Either vehicle can be employed to insert the content into working memory.

Extended mind grants that there are two kinds of cognitive case that extend beyond the skull. First, our *dispositions to behave* partly derive from external support for dispositional knowledge. This is a claim about engineering information into our environments in order that brains may function more efficiently and accurately. This is niche construction. Clark and Chalmers give the example of Otto and his notebook. Otto keeps information in his notebook, such as the location of the Museum of Modern Art. When asked how to get to the Museum of Modern Art, Otto is disposed to produce the right directions because of the external, readily available representations in his notebook. A more interesting case is found in Clark (2003). Alzheimer's patients with severe disease are often found managing to cope still living alone. This is because they engineer their environments to prompt them. They keep utensils in open view, or write post-it notes to themselves. This structured environment disposes them to behave in certain ways. When admitted to hospital, these patients deteriorate noticeably. It is like they have suffered a sudden further bout of brain damage. Which in a very real sense they have. They have lost their external cognitive scaffolds.

The second way in which cognition allegedly bleeds into the environment is more interesting. This is the claim of *active information processing loops*. We utilise active loops of information processing into the environment (of non-biological media) and

this constitutes proper parts of our cognitive systems. These two phenomena work together and constitute the extended mind.

For many the concept of the extended mind rests upon a *parity principle*. This can be formulated in several ways but the basic sense is that a process counts as cognitive if it is the case that had that process occurred in the brain then one would have no hesitation in attributing cognition to it. Importantly, others, such as Sutton et al. (2010) and Sterelny (2010), argue more for the *complementarity* relation I set out for hybrid thoughts in the previous section, than for parity. Sutton sets out three options:

(1) There is true extended cognition where cognitive processes are sometimes constituted by external processes (perhaps based on parity).
(2) There is merely embedded cognition where cognition causally interacts with external resources but remains intracranial (cognition is never actually hybrid, it is merely supported by some external factors).
(3) Distributed or scaffolded cognition where substantial and surprising interactions occur between internal and external resources such that cognitive psychology should study these distributed processes.

Sutton's project is to develop option three, which neither requires a cognitive extension thesis (1), nor completely denies that there are such extended systems. Rather than parity, Sutton et al. favour the notion of *complementarity*. Biological resources and external resources don't need to have the same formats or dynamics to function in integrated systems that enhance cognitive abilities. Parity or functional isomorphism is not required for interesting integrated systems. Clark himself argues in a similar vein:

> 'The argument for the extended mind thus turns primarily on the way disparate inner and outer components may co-operate so as to yield integrated larger systems capable of supporting various (often quite advanced) forms of adaptive success' (1998, p. 99).

Complementarity also knits with the approaches of theorists in cognitive anthropology such as Merlin Donald's work discussed above. Finally,

complementarity does not require that artifacts or external resources have mental states, or memories, or experiences, or are otherwise 'cognitive', it is simply that artifacts cooperate and coordinate with quite disparate internal and external resources to constitute systems that process information. The complementarity argument can be seen to subsume the parity argument. All cases of parity are cases of complimentary relations between inner and outer resources.

If we compare Otto's notebook and the hospital notes of some patient, then we find some important differences. Otto refers to his notebook expecting the memories he has recorded there to remain the same. They are durable and reliable. A doctor may refer to the hospital notes to follow sequential blood test results; she expects the notes to change, they are dynamic. So we see that some cognitive artifacts are personalized for one user, this may involve bookmarks, annotations, or additions to enhance the ease of use. Repeat users may come to depend on wear and tear patterns of the extended resources they employ. Other cognitive artifacts are intended for group use. Sterelny (2010) takes this to explain why some external resources may be seen more appropriately as extensions of the (individual) mind, whereas other (perhaps many more) artifacts are merely resources in extended systems of information processing. Sterelny thinks little actually hinges on the distinction between cognitive extension and mere supported or scaffolded extension. Hence the extended mind picture is not false, it is merely not the most useful way to approach these more disparate phenomena. I reiterate again the two scaffold relations. First is the temporary kind of scaffolding (a), which allows something to be built, or allows brains to develop a particular way. This kind of scaffolding is not always required for later function. Second are scaffolded processes (b), which, like a vine growing up a trellis, always require the scaffolds in place in order to function.

It is not entirely clear how (a) and (b) fit into Sutton et al.'s triple framework above. They don't seem to fall under (1), just as the scaffolding is not a constitutive part of the building. Also, cognition could (2) 'remain intracranial' whilst being scaffolded in operation, or requiring scaffolding by external resources for development. And clearly real time scaffolded processes (3) ought to be studied by cognitive science.

In the context of the present thesis, I hope that it is clear that all three of Sutton et al.'s situations are *causal* of phenotypes, when we consider psychological traits to be capabilities, and should be considered important in the development and evolution of these traits. For the purposes of the project in this thesis I can be neutral between (1), (2), and (3) in Sutton et al.'s schema. Though I favour options (1) and (3) over a strict approach denying interestingly extended and distributed systems for reasons that should be clear from reading this chapter. It is the *causal* dependence of psychological traits on external technologies that is most of interest to me, not whether minds are ontologically extended into the world. However, discussing in detail the possibility of ontological extension highlights very clearly this causal dependence.

What this debate highlights is that the constitutive elements of processes can certainly be heterogeneous in nature and coupled to each other in various and interesting ways. There are also processes that are shared, multi-agent processes. People can collaborate to perform useful, information gathering and information manipulating functions in the world. Such things as conducting a census, or processing a patient for surgery fit these categories. Sometimes the multi-agent process may perform a function that we would ordinarily label cognitive, such as working out the solution to a complex mathematical problem. The process that has the function of extracting oil from the ground is constituted by a heterogeneous set of causal elements. These include human labourers, industrial drills, static pipes, complex mechanical pumps, and so on. A mix of biological and non-biological components performs the task. Whether cognitive processes are constituted by symbols, pens and paper, hand gestures, or external lists, there is a causal dependence of human behaviour on this wealth of external material. It should be clear that we could change the external materials and thereby enable or disable a host of different individual or social human functional traits. In particular we could change external materials and enable or disable individual or social human psychological traits. I think that some examples, such as the long multiplication example, have the right characteristics to count as cognitive on a sensible account of

what is cognitive. Hence some legitimate cases of 'extended mind' in the strong sense exist. But I concede that a lot of informational engineering of our environments doesn't have an 'owner' in the right sort of way to count as cognitive, yet this sort of engineering can still be very much *causal* of psychological traits and capabilities. I also claim that many such traits are illuminatingly hybrid. This hybridity seems to mark human psychological traits as very often different in quality from non-human psychological traits.

## Evolution

Remember that niche construction activities can also drive evolution. I now want to move from discussions of the development and function of psychological traits to some considerations for evolution. If cognition is embedded in technology then it will be sensitive to technological change and variation. Descent with modification permits the inheritance and accumulation of culture. Writing systems may be learned by subsequent generations and then altered or added to, thus we can get 'ratchet effects' in culture and rapid accumulation of novel cultural items (Tomasello 1999). But writing systems also become a part of the developmental environment of subsequent generations. Children born into a world partly constituted by accumulating and somewhat diverse writings experience a novel developmental context in each generation. It seems likely that the very presence of some primitive signs and writings, which arose through variation, inheritance and natural selection, also provided the basis for a cognitive developmental ratchet. I discuss this in chapter eight.

Such ratchet effects are worth remembering when trying to understand whether the origins of writing led to dramatic social, economic and cognitive change, or whether writing systems were the result of such change (Barton & Hamilton 1996, p. 808). The answer is probably both. Writing systems arising drive changes in the developmental context, which then drives further elaboration of writing and social changes, thus providing heritable variation for selection to act upon.

Psychological traits can change in various ways. They can change because brains change or they can change because the ways that brains relate to the body or external world change. Change in internal brain architecture is derived from changes in development. This could be genetic change, environmental change, or cultural contextual change, for example, as I have argued in chapters one through three. From an evolutionary point of view any of these changes, such as technological innovation, could drive changes in the distributions of psychological traits in populations over time. Both Donald (1991) and Clark (2008) appreciate this.

Furthermore, in any actual cases of extended mind, external objects do not just cause derived developmental changes, they also constitute parts of cognitive processes. The external part would be a cognitive component independently of any developmental effect it has. So external objects (technologies) can have developmental effects, can support functional activities, or possibly could constitute parts of cognitive systems. Cognition results from the interaction of brain and world (see figure 6).

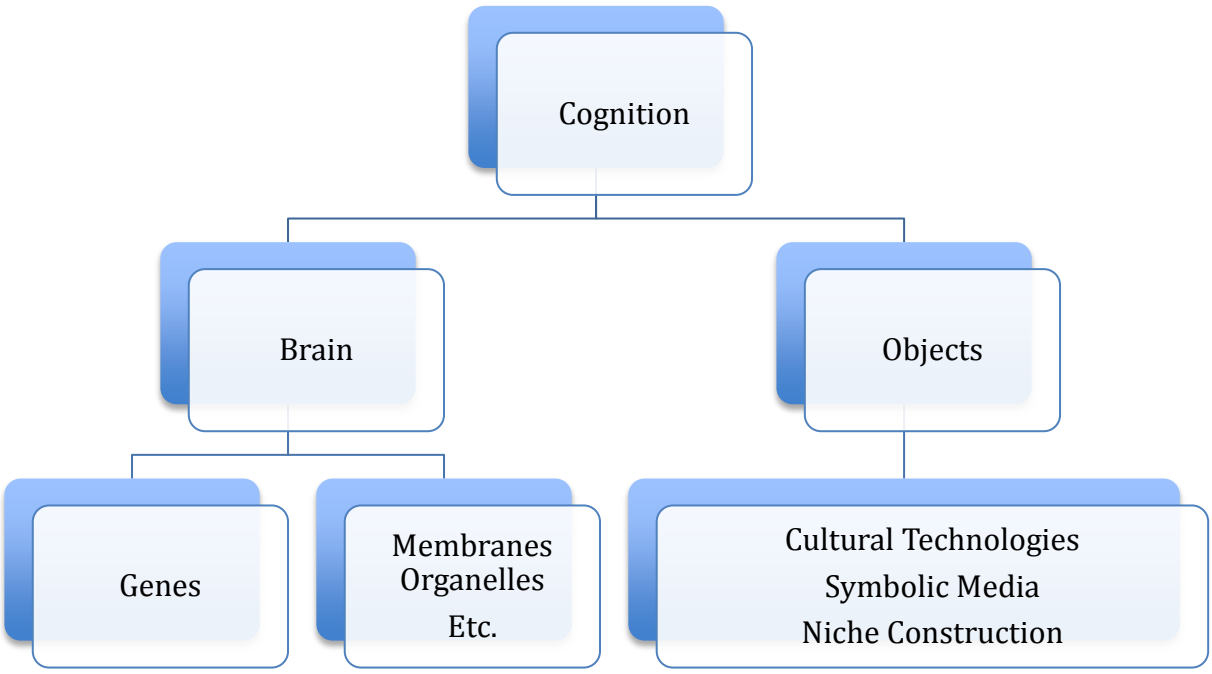


Figure 6: Constitutive Components of Cognitive Systems (for simplicity I have omitted the body, which though important, was less discussed in this chapter). Alterations to any of the base causes of brains, or objects, e.g. genes, organelles, technology, etc, could alter cognition.

## Summary

Chapters one through three have shown us that the development and function of a brain is importantly derived from external contexts. We couldn't do mathematics, or write chapters of syllogistic thought without scaffolding and/or extending technology. What humans can do has changed dramatically over time. Psychological traits as I understand them have changed.

The developmental context of *Homo sapiens* has been evolving as our culture evolves. There will also have been *some* genetic change constituting changes in brains in the last 60,000 years. But importantly, there has been a massive amount of change and evolution of our material worlds. This has provided fresh opportunities for brains to engage in complimentary fashion with the external environment to support a range of complex psychological functions. There has been considerable recent evolution of these traits, driven largely by changes in our material context and the relations of brains to external objects and symbol systems. From an explanatory point of view the causal dependence of function on external supports is sufficient. Arguments claiming the constitutive dependence of psychological functions on external materials serve to further highlight this importance.

The point that I want to emphasize is that there seems to be no real boundary between the neural and the external, the causal chains pass seamlessly from one domain to the other and back again. Rather than being fixated on what is a psychological trait, or what constitutes the mind, perhaps we ought to be more focussed on human capabilities. What can an agent achieve in the world and what factors are causally important in that achievement? What constitutes the system that allows an agent to perform the tasks that they do?

With this background (chapters one through three) in mind, we can enter the main argument of this thesis. In chapters four through seven, I will develop and employ principled reasons to explain the privileged role that human technology plays in psychological development. In chapter eight I will return to the issue of the evolution of human capabilities, and will examine two problems of prehistory, the sapient paradox and the Neanderthal extinction. The first step of all this is to turn, through chapters four and five, to an analysis of what it is to be a causal component of a developing system, and to look at how different genetic and non-genetic (including technological) elements relate to each other and to their effects.

# PART TWO:

## Untangling Causation in Development

# CHAPTER FOUR – Causation in Biology and Development

## Introduction

To this point we have seen that there are a range of cultural technologies that have causal effects on psychological development. In chapter two I showed that these influences are modulated by a highly plastic neural system. Finally, we've seen that the functioning, and in some cases constitution, of the mind is causally dependent on a range of technological objects in a changing matrix external to the body itself. I turn now to ask, given that there are a great many influences on this system (genetic, environmental, cultural) how do we best understand the causal relations that hold between all these disparate elements? If technological culture is *a* cause of cognition, what *kind* of cause is it? And how does cultural causation compare to, say, genetic causation? If we want to see how technology fits into the broader picture of cognitive development and cognitive evolution then we will need some principled way of determining what elements play what roles in development. In other words, what is important, when, and why? In this chapter I want to introduce a tool for answering such questions.

The motivation for such a tool is threefold. First is the need to untangle the complex picture of cultural technologies and cognitive development that I have thus far explicated. Are all the technological determinants I have described equal, or similar in their causal profiles? It would be nice to be able to bring some structure to this picture. Second, such structure may help quell lingering adherence to genetic determinism, in one form or another, which is still held by some investigators. Genes demonstrably do not determine phenotypes, and do not even play a homogenous causal role when it comes to development. The model I propose will clearly elucidate this fact. Finally, developmental systems theory has appeared as an awkward challenge to any attempt to build simple models of developmental causation, or any attempt to favour some developmental causes over others. I hope that the model I

propose can bridge the gap between overly fuzzy, holistic developmental systems approaches and overly simplistic genetic (or cultural) determinist ones. To be clear, the project I begin in this chapter is one of explanation. I am not interested in the metaphysics of causation, or what it means for A to cause B. I am taking for granted that elements of the developmental matrix do cause traits. I am trying to explain the emergence of traits with some degree of generality. For psychological traits, some part of this explanation will be cultural and technological. Furthermore, although I employ mainly examples of biological cause and effect in this chapter, I believe that the general strategy I employ ought to work for all cause and effect relations.

In this chapter I will outline the gene-environment interactionist consensus and developmental systems theory. I will explain three forms of the 'parity thesis' advanced by developmental systems theorists[7]. I will show that only one of these need be taken seriously if the project of disentangling developmental causation is to work. I will then discuss some reasons why one might be tempted to cling to a form of genetic favouritism, but I will dismiss this approach by arguing against the informational nature of the genome with respect to developmental processes. Then I outline some various approaches to causal analysis and causal modelling. I articulate what it is that we want from a causal analysis and suggest that James Woodward's approach gives us a likelihood of success. I will then identify a key lingering problem with Woodward's approach and work out a fix.

## The Interactionist Consensus

The interactionist consensus (Sterelny & Griffiths 1999, Pennington et al. 2009) describes the universal acceptance that all biological (psychological) traits develop as a result of the interaction of genetic and non-genetic factors. This comes as no surprise once we understand the arguments in chapter two, which described how brains and psychological traits are constructed contextually.

[7] This developmental resources parity thesis is not to be confused with the parity arguments advanced in the last chapter, which claim that some external cognitive representational resources are on a par with internal neural resources and that hence we have an 'extended mind'.

Neither genes nor environments alone determine the final phenotype of an organism. Genes and environments interact to produce phenotypes. We've seen this already with arguments emphasizing both the heritability and contextual dependence of human IQ. Clearly neither genes nor environments alone can produce an organism. Indeed, *just* genes and environments are still insufficient, because we need such things as membranes and other cellular resources to be present as well. Such resources seem to be a bit of a gray zone, neither genes, nor environments.

Although to this point I have largely focused on plasticity responses to the environment, genes have a role to play in development. A host of interesting gene-environment (G x E) interactions exist and most likely apply to all aspects of cognition. These effects are known as resilience, diathesis stress, and bio-ecological effects among others. I now give a few examples. For more examples see Pennington et al. (2009).

Resilience effects are developmental outcomes that occur despite having a favourable or unfavourable environment (or genome) present. For example, it is known that the heritability of good reading is higher if the parents' level of education is low. The child tends toward good reading, despite the poor parental education. Resilience would be universal if some form of determinism (environmental or genetic) were true.

Diathesis stress reactions occur when the agent possesses some risk factor, for example a genetic or environmental risk factor. This risk factor is then stressed (by either the environment or genes) and the effect, which was at risk of occurring, manifests. For example, it is the case with psychopathologies, such as conduct disorder and depression, that environmental stress coupled with genetic risk factors results in more disordered behaviour than would otherwise be expected by either factor alone or in additive combination. Conversely it is the case that someone could possess genetic risk factors for major depression, but not ever manifest depression

because the genetic risk factors are never coupled with the appropriate environmental stress.

Bio-ecological effects are where certain environments will permit underlying genetic differences among agents to show themselves. Lewontin's (1995) example of genetically variable seeds demonstrates this. When planted in poor soil all the seeds fail to thrive. When planted in rich soil the seeds thrive according to their genetic predispositions. Hence, a favourable environment actually exaggerates the genetic differences. I have indicated something similar in figure 3 (chapter one) when discussing human IQ. As I mentioned, it is reasonable to assume that such effects are likely to play a role in the development of all cognitive phenotypes. Note also that none of this requires a coding or informational relation between the genes (or environments) and the phenotypes. These are merely causal effects that some genes have when interacting with some environments.

## Developmental Systems Theory

Developmental systems theory (e.g. Oyama, 1985; Oyama, 2000; Oyama, Griffiths & Gray, 2001) seizes upon this interactionist framework and argues that there is no *a priori* reason to privilege some developmental causes (notably genes) over others. Importantly, developmental systems theorists don't always deny that there *could* be privileged causes. It is just that it is not obvious what these might be and they don't always think it will be the same kinds of causes in all cases. But most importantly, developmental systems theory proponents hold that to look at the interaction between genes and environments through the lens of G x E interactions is to hold on to a false conceptual dichotomy. In fact, there is a very real sense that the heart of developmental systems theory is opposition to nature/nurture, or genes/environment dichotomies (Stotz 2010).

Developmental systems theory is also an evolutionary thesis. It is important to realize that there are two key processes going on. One is the extra-genetic component of the developmental system interacting with the genetic component during ontogeny. The

other is the possibility of inheriting these extra-genetic components from generation to generation. Developmental systems theory emphasizes the relevance of development to evolution and emphasizes the evolutionary potential of extra-genetic inheritance. The fundamental replicator therefore is the 'life-cycle' generated through interaction of a developing organism with its environment (Griffiths & Gray 2005). All sorts of things are inherited not just genes.

Developmental systems theorists often claim that what changes over evolutionary time is a 'developmental system' constituted by the organism and some broader developmental context. Many aspects of 'environment' are taken seriously as products of evolution in contemporary biological theory. For examples we can look to *niche construction*, which emphasizes the environment-altering activities of organisms (Odling-Smee, Laland, Feldman 2003), *dual-inheritance*, focusing on cultural as well as genetic inheritance (Richerson & Boyd 2005), and *parallel hereditary systems* including genetic, epigenetic, behavioural and symbolic channels (Jablonka & Lamb 2005a). Culture and cultural technologies are obviously a key part of this sort of system for humans. Developmental systems theory claims that any resource that is reliably present in successive generations is inherited. Genes are not special on the basis of inheritance.

So how ought we to approach these developmental systems? Oyama suggests that we should entertain 'parity of reasoning' with respect to the different developmental causes. In addition we must circumscribe the *relevant* causes in any given case, for otherwise analysis or manipulation will become impossible. This is why Godfrey-Smith (2001) cautions against insisting that we take every causal factor equally seriously.

## The Parity Theses

Difference makers are events that make a difference to some outcome. Had the event not occurred, then some effect would not have occurred, or would have occurred differently. From chapters 1-3 we can see that cultures can be difference makers in cognitive development. But, do any elements of culture exhibit symmetry

with the causal role of genes in development? Or do genes have some causal primacy? Developmental systems reasoning leads us to the *Parity Thesis*. No distinctions between developmental elements map neatly on to DNA on the one hand and all other causal factors on the other, therefore causal privileging of the genome is unwarranted.

There are actually at least three parity theses:

(1) Genes and *some* environmental causes are on a par in development. One could argue that given genotype A, only by knowing environmental cause B, could we know something about the phenotype. Or given environmental cause B, then we need to know phenotype C in order to know about the genotype. In a sense genes and some environmental causes would share an explanatory symmetry. This is the sort of thing that people who say, 'Your genes *and* your upbringing are equally important in how you turn out' probably mean. Here's a concrete example. In the metabolic disease phenylketonuria people with a defective gene and phenylalanine in their diet produce urine with a distinctive odor due to a build up of phenylalanine in the body. If we know that the subject has the defective gene, then only by knowing if there is phenylalanine in their diet could we know whether their urine will have a distinctive odor. Or given that we know that there is phenylalanine in the diet, only by also knowing the odor of the urine, could we infer anything about the genotype. In this particular case the defective gene and the contents of the diet seem to be on a par.

Sometimes a second, stronger, Brute Parity Thesis could be, a little uncharitably, reconstructed from some DST writings. I articulate it here in order to fully map out the conceptual space of positions on parity.

(2) This is the idea that *all* elements of the developmental matrix are equally causally privileged. Membranes are just as important as the temperature of the womb[8], and both are just as important as genotypes and social upbringings. This is the strongest

[8] The incubation temperature of turtle eggs determines the sex of the offspring, the offspring's chromosomes do not.

sense of the developmental systems argument and the most awkward for any hope of explanatory traction. In this version of the parity thesis everything that is a difference maker is equal. Many factors are necessary and none is sufficient for developmental effects.

Finally, a Parity of Reasoning Thesis may be meant.

(3) This is the idea that each and every element of the developmental system must be approached *as if* it might be important in development and we can take things from there. This means that everything is a *candidate* for importance. We cannot on general theoretical grounds rule any factor in or out. The tool I explain later in this chapter will *provide principled grounds* for ruling causal factors in and out as explanatorily important.

Taking the third parity thesis as our benchmark for analysis we will discover whether either of the other theses is warranted, but not vice versa. So this is my preferred strategy. In the causal analysis that follows, no causal element will receive treatment any different from any other. All developmental causes are candidates for inclusion in the causal explanations. However, it will then become apparent that not all developmental causes are equal, and we will be able to start selecting those causes that must be included in useful explanatory models.

## Genes and Causal Privilege

Causal privilege is the idea that in a complex causal system we want to focus our attention on the 'important' causes. This is particularly so when trying to intervene on a system. By showing asymmetries among causes we can favour some causes over others according to the questions we are asking. These favoured causes (arguably genes in the present case) are the privileged causes. Furthermore, privileged causes should tend to explain the variations across systems (e.g. see Russo 2009).

In the interactionist view described above genes are, 'context-sensitive difference makers' in development (Sterelny & Griffiths 1999, p. 99). Could they be more than this, and on what basis? Maynard Smith (2000) argues that genes carry semantic information about phenotypes and that this fact makes them specially privileged causes.

Sterelny and Godfrey-Smith (2007) identify four possible applications of informational concepts in biology. One of these is the description of whole organism phenotypic traits as specified or coded for by information contained in the genes. This seems to be the sense of information most commonly applied to the genome. We sometimes speak of this or that gene being a gene for (semantic content) some phenotypic trait. The other important possible application is the treatment of cellular processes as executing some sort of program. It is possible that genes, as well as (instead of) representing phenotypic outcomes, have instructional information content, that is, they may be analogous to commands or imperatives. I will discuss genes as programs below.

I now offer three arguments *against* the informational role of the gene, and therefore reject developmental causal privilege of the genome on these grounds. This will return us to the parity thesis as our starting point.

## Information and the Genome – The Great Disappearing Act

Supporters of developmental systems theory hold that there is no dichotomy to be drawn between genes on the one hand and environments on the other. There is no *a priori* reason to privilege genes in any developmental cascade and they must be approached on equal merit with all other elements of the system. However, in biological science we often talk of genes as 'coding', 'representing', 'transmitting', or 'instructing'. Perhaps we can argue that genes are to be privileged due to the information content that they hold. I wish to quell such appeals by arguing that it is at best highly controversial that there is any relevant information content in genes with respect to developmental outcomes. Therefore, we ought to avoid talking about

genes in such ways. I will give three arguments against the existence of information content in the genome and conclude that, therefore, we ought to stop treating the genome as an informational reservoir with respect to *development*. This does not necessarily rule out information content in genes with respect to *inheritance* (an approach favoured by Bergstrom & Rosvall 2009, and Shea 2006). But this is a separate issue.

## What is Information?

We must be clear about the relevant sense of information with respect to the privileging of genomes in development. I will explain four approaches, the mathematical sense of information, the natural sign sense of information, the transmission sense of information, and the semantic sense of information. Only carrying information in the last sense could privilege genes in developmental systems.

Shannon (1948) formalized the *mathematical sense of information*. According to his theory information is transmitted from a source (via coding), to a receiver (which decodes), through a channel. The easiest way to conceptualize this is to imagine a television transmitter, sending electromagnetic waves through the atmosphere to a television set. According to Shannon's theory the number of different messages that could have been sent sets the *amount* of information that can be conveyed in a particular channel. Information theory in this sense defines information as the probability of a particular message being selected from the set of all possible messages. Shannon's is a theory of the quantity of information that exists. It says nothing about the content of a signal.

If we conceptualize the genome as a source, the extant physiological conditions as the channel, and perhaps the cell as a receiver, we could quantify how much information could exist in the genome. Shannon demonstrated that because every form of information can be encoded in bits, there is an essential unity of every mode of communication (telephone, text, radio, pictures). It is possible to conceptualize the genome in this way too. However, what this approach misses entirely is the content

of the information, the semantics of what is being conveyed (Perez-Montoro 2007). This point is emphasized by Jablonka & Lamb (2005b) who note that a DNA sequence coding for a fully functioning enzyme contains the same amount of information as a nonsense string of DNA, even though the second string is semantically 'meaningless'. It is this semantic notion of information that is usually being employed when we are discussing the informational content of genomes. We want to know what the genome is saying. Not how much it is saying.

Grice (1957) argues that whenever Y is correlated with X we can say it carries information about X. This is the basis of the *natural signs* account of information. A natural sign conveys information that something is the case in the world. For example lightning conveys the information that there is a storm, rabbit footprints convey the information that a rabbit has passed, smoke conveys information about fire, tree rings about the age of a tree, and so on. Natural signs are effects of physical processes and they give information about physical states of affairs in the world. This is why a natural signs account could also be called a causal account of information. But this account of information gives no account of the direction of information flow. Smoke is correlated with fire, and fire is correlated with smoke. Mathematically, knowing a phenotype gives you the same amount of information about the genotype as knowing the genotype gives you about the phenotype (Griffiths and Gray 1994).

All elements of the developmental system are natural signs. Membranes, transfer-RNAs, the temperature of the environment, all these give us correlational information. However, different forms of membranes aren't natural signs of phenotype *difference* in the way that genes are. Cell membranes correlate with all phenotypes, but different genes correlate with different phenotypes to different degrees. In this sense genes are correlated (however loosely) with phenotypes and hence carry information about them. But then so too do many other elements of the developmental matrix, such as an infants diet, or the existence of number words in a language. It certainly seems like natural signs carry information about states of affairs, but this 'aboutness' is not representation it is correlation. This 'aboutness' is simply due to natural signs being the effects of past causes. Genes may correlate

more closely with some phenotypes than membranes or temperatures do, but it is not obvious that they correlate more closely to certain cognitive phenotypes than number words do. Genes are at best natural signs, but so are many elements of the developmental matrix.

Let's move on to the *semantic view of information*. When I draw a sketch of a dog on a piece of paper there are at least three different informational ways of analyzing this:

1. The number of bits of information, the quantity of information in the picture. This is the mathematical sense of information.
2. As a natural sign that someone drew on the page. This is the causal view of information.
3. As having the semantic content {DOG}, i.e. being *about* a dog, or *representing* a dog.

The semantic approach captures the directionality of representational content. The sketch represents the dog, but the dog doesn't represent the sketch. This seems to be important when we are looking at the relationship between genes and phenotypes. Genes are sometimes argued to represent phenotypes, but phenotypes aren't usually argued to represent genes. Furthermore, those who favour the semantic view of genetic information (e.g. Maynard Smith 2000) tend to think that genes represent phenotypes, but that other elements of the developmental matrix, such as environmental factors, do not. I emphasize though that the semantic approach is still susceptible to parity arguments. There could be other kinds of developmental resources than genes, which represent phenotypes. So although carrying semantic content could be a reason to privilege genes, this may not privilege only genes. But we will first see whether or not we are justified in attributing semantic content about developmental outcomes to genes. I'll return to this question shortly.

The final sense of information that it is important to address is Bergstrom & Rosvall's (2009) *transmission sense of information*. These authors argue in favour of an important informational role for genes. But they restrict this role to cross-

generational transmission of information, not a developmental role for genetic information. They note that information theory is a theory about the efficient transmission of information. Information is packaged and sent across space or time. Genetic transmission is about transmission of a genotype from generation 1 to generation 2, not from genotype to phenotype. Their transmission sense of information is defined as follows:

> 'An object X conveys information if the function of X is to reduce, by virtue of its sequence properties, uncertainty on the part of an agent who observes X.'
> (Bergstrom & Rosvall, 2009, p. 6, online version)

Examining the sequence properties of genes we see that they are built for transmitting information. When we look cross-generationally parity arguments are 'shattered' (p. 7, online version) because no other elements of the developmental matrix seem to be designed (selected) to carry information as efficiently and with such a capacity as genes. The very structure of genes suggests that they are *for* the transmission of information across generations. This may be so, but this does not mean that there is any semantic, or transmission sense, information available in developmental (as opposed to phylogenetic) processes.

Bergstrom and Rosvall's arguments for a 'transmission' sense of information give privilege to genes in evolutionary processes. Sterelny (2001) and Sterelny in Sterelny & Griffiths (1999) also argue that replicators are to be privileged in such explanations. However, these approaches do not explain any *developmental* causal privilege possessed by genes, and these approaches (as Sterelny is well aware) are susceptible to parity arguments, which suggest that other extra-genetic factors can fill the same role (see Jablonka & Lamb 2005a). The hereditary (vertical) conception of informational genes can be seen as a separate issue to the 'horizontal' transmission of information.

Clearly there is Shannon information in the genome, but there is Shannon information everywhere in the developmental system. Just as there is information in the electromagnetic signal correlating with the TV picture, there is information in the

TV picture correlating with the atmospheric conditions. This universality and reciprocity of information is uninteresting in the present context. Clearly genes are also natural signs. Any natural sign is just a necessary result of a physical phenomenon. To the right interpreter genes are certainly natural signs that, for example, some parental organism with such and such a genome survived to reproduce. Genes are also natural signs of phenotypes, but only in broad correlational strokes. If we want to attribute some sort of privileged causal responsibility to genes we are after more than mere smoke-and-fire relations.

### Denial of Genetic Information (1): Syntactic Programs

Perhaps there is a genetic program. That is, perhaps genes are instantiating a set of rules, a recipe, or instructions that leads them to produce a certain set of phenotypes and somewhere in this system resides information, or, more particularly, information about phenotypes. Perhaps on this basis we can privilege genes in developmental explanations.

We can see how a physical system can instantiate quite complex logic without bearing semantic information. Imagine an empty glass is floating upright in a bath. Two other half-full glasses, A and B, can be poured into the first glass. If one is poured in the first glass contiues to float, if both are poured in then the first glass sinks. A and B and the floating glass constitute a logical AND operation. If and only if A AND B are poured will the glass sink. So whether the glass is afloat or not represents the truth of the proposition A AND B. But the floating glass is not doing logic, we just have a regularity of causal specificity, no more, no less. *We* can *interpret* the system as a logical AND gate, but the system itself is no such thing. One might argue that the same applies with any appeal to the genome as a series of imperative IF-THEN statements. The genome is merely a physical system obeying rules of causal specificity. However, this is too quick. In the bath case nothing in the system is designed to respond to the glass sinking. In the case of an organism, there may be some response that is adaptive only if A and B are both true. If there is selection for the mechanism that activates if A and B are both true, then over time we may

legitimately claim that the mechanism (perhaps some gene) has the function of testing whether A and B is true.

However, even if the genome instantiates such a program, then this does not mean that it bears informational content in a way required to differentiate it from other factors in development. More specifically, it doesn't mean that the genome bears semantic content about phenotypic outcomes. Rosenberg (2006) denies the informational content of genes on precisely the grounds that at best the genes are instantiating a program. He adapts Searle's (1980) Chinese Room argument to the case of the genetic 'program'. The Chinese Room argument is a response to the research program of strong artificial intelligence (AI). Strong AI claims that if we can uncover the program that human thought instantiates in the brain, then we will understand the brain. Importantly, human thought is 'about' something. It has semantic content and is intentional. In the Chinese Room thought experiment, a person is manipulating unfamiliar symbols according to the instructions in a book in order to produce intelligent utterances. However, the person does not understand the symbols as content, they merely have syntactic form. The person in the room is instantiating a rule governed program, and producing coherent outputs, but does not understand anything that passes through their hands. To the person in the room there merely exist contentless symbols. Hence, instantiating a program is not sufficient for there to be semantic content. The genome, too, possesses syntax. There are right and wrong ways for nucleotides to be aligned, there are rules (a code) mapping codons to amino acids to produce outcomes. However, nothing in this system of rules understands the 'content' of the molecular symbols. But, just because instatntiating a program is not sufficient for semantic content does not mean, yet, that there is no such content in the genome. The genome could be both instantiating a program *and* possess semantic content.

Rosenberg's reply is that there is no original intentionality contained in the symbols (or by analogy the genome) as they depend on an external observer, who understands them, in order to possess meaning. Rosenberg argues that even if there is true arbitrariness in the genome, then the genome's lack of original intentionality

(all meaning is derived from us, the observers) rules out any special informational status it might be said to have. Rosenberg claims that a useful test of whether a state is intentional, or if a state is 'about' another thing, or represents another thing, or has informational content, is that the truth or falsity of a description of the representing state ought to be sensitive to how the content is described. Take the example of a mental state. Rosenberg uses the following example. Lois Lane believes that Superman was born on the planet Krypton. Let's assume this is true. However, Superman under a different description is Clark Kent. But presumably Lois Lane doesn't believe that Clark Kent was born on Krypton. By making a truth-preserving substitution in the description we have changed the truth value of the content of Lois Lane's belief. This, argues Rosenberg, is paradigmatic of intentional content. He then argues that this feature is clearly lacking with respect to genes. For example, when we say that the triplet sequence CAT represents the amino acid histidine, or is 'about' histidine, then there ought to be some way that we can describe histidine that makes the statement 'CAT means histidine' or 'CAT is about histidine' false. There is no such description. CAT means the only amino acid spelled with an initial 'h' in English, CAT's content is about Francis Crick's favourite molecule, and so on, are all true. Hence, argues Rosenberg, genes are not intentional (do not possess semantic content).

One concern about this view of Rosenberg's is that he might be conflating information with intentionality. There being different kinds of information, some of which are intentional representations and some of which are not. The real question is what sort of information we require of the genome in order to attribute a special informational role to it with respect to development? I've already argued that the quality of being Shannon information or of being a natural sign is insufficient. It seems that the relevant kind of information must be semantic and representational. But even if genes are instantiating a program, then they need not be informational in the right way.

So, it might be thought that the genome could be instantiating a program, but that this is not sufficient to grant it semantic content about phenotypes. This is Rosenberg's view. It might be further thought that instantiating a program is

insufficient to grant developmental causal privilege because this just means that any aspect of the developmental matrix that is instantiating a program should be privileged. Perhaps cell-division is instantiation of a program, perhaps the creation of antibodies to fight infection is instantiating a program, perhaps mothers teaching infants to read is instantiating a program. This is not sufficient to privilege the genome over other elements. This is how developmental systems theorists ought to reply to Rosenberg-like arguments. However, if the genome bore semantic information about phenotypes, or represented the outcomes of development this may be a different story.

### Denial of Semantic Information (2): Genes do not represent developmental outcomes

A weather report clearly represents the next day's weather. It bears some relationship with the actual weather that has the possibility of being more or less accurate, just as a sketch of a dog represents the dog more or less accurately. The reason that a weather report (or an architect's plan for that matter) is a representation is that that is its intended purpose. An ant trail in the sand (no matter how much it looks like the Virgin Mary) doesn't *represent* anything because its function is not to represent. The ant trail, of course, carries *information* about the path the ant took. The information in this case is a true proposition. It is sometimes argued that genes represent phenotypes (e.g. Sterelny, Smith and Dickison, 1996). However, if the ant trail doesn't represent the Virgin Mary because that is not its function, then we must ask how do genes come to possess the function of representing?

Millikan (1984) provides theories of biosemantics and teleosemantics to account for representation in biology. Representations or semantic content occur in biology where there has been selection *for the function* of representing. For example, some bacteria possess a tiny inner magnet which causes the bacteria to head toward the magnetic north pole (and hence head down to less oxygen-rich water). Intuitively the pull of the magnetosome represents the location of oxygen-free water. This is the magnetosome's function. It has been selected over evolutionary time for this

function. The magnetosome does not represent the location of the magnetic north pole, because this is not what its consumers require it to represent. The *consumers* of the representation require only that it represents oxygen-free water.

The representing entity need not represent correctly or accurately all the time, for example the magnetosome can be disrupted by placing a bar magnet near the bacteria. The representing entity need only correspond to the represented content *often enough* to establish a selective history for this function. This approach could account for genes misrepresenting phenotypes in different contexts. Millikan emphasizes the need for both producers and consumers of representations. It is the producers' function to produce a representation that indicates something, and the representation *indicates something to a consumer* of representations. Consumers alter their behaviour according to the information in the representation. The producer's function is to produce correct representations, but if conditions are unfavourable then it will fail. Being reliable always depends on external circumstances. What counts as 'too unreliable' is a function of the costs incurred when representations turn out to be false versus the gains made when they are true.

Similarly, Maynard-Smith (2000) argues that there are indeed coders, transmitters, receivers, decoders and information channels through which the genome can be seen as the bearer of information. He argues that in the developmental matrix only genes have this privileged informational role, and that the reason they do is because the information got there when the process of natural selection 'put' the information there. The information is required so that rich adaptive complexity can arise afresh in each generation. According to Maynard Smith's argument rich semantic properties exist in biology where there has been the right kind of history of natural selection. Genes carry information about phenotypes because that is their function in evolved organisms. Although Millikan and Maynard Smith claim to identify interpreters (or consumers) in biological systems I am sceptical. I give a defense of my position in the section titled 'no interpreters' below.

However there is a further worry for the 'selective history builds representational content' approach of Millikan and Maynard Smith. The problem arises when we consider that representing phenotypes, and representing developmental outcomes are not the same thing. Shea (2006) discusses representation in the genome and notes that there are two ways of answering the question of how adaptive complexity arises in each new generation, one is ontogenetic, and one is phylogenetic. Shea argues that only the second case requires representation. All genes have developmental properties [genes-D] but some genes also have correlational or predictive properties [genes-P]. The presence of a gene-P correlates with some phenotypic property. Individual genes will contribute to many developmental outcomes, but individual genes persist from generation to generation because *some* of these outcomes are selected for, hence selecting for the gene. Genes-P properties arise when genes are selected because of their tendency to bring about some phenotype through their gene-D activities. Gene-P properties are selectional properties and give rise to teleofunctions. The teleofunctions of genes do not feature in ontogenetic explanation. Put another way, gene-P properties do not figure in developmental explanations and cannot be the basis of causal privilege in development. Genes-D explain stage by stage causal processes in development. Developmental explanations are complete without appealing to gene-P properties. Genes-P properties are phylogenetic representational properties. So even if it is true that genes represent phenotypes due to selective history, this does not mean that we ought to privilege genes in developmental explanations. Furthermore, Shea argues that DNA codes for amino acids, by virtue of a robust mapping rule from triplet codons to protein chains, but being a code and being a representation are two separate issues. A spy could employ a code to map between an encrypted string of letters and some decrypted string of letters. This is so whether or not the strings actually represent anything. So genes-D properties and the existence of a genetic code fully account for genes' role in development, without positing any representation of the outcomes of developmental processes.

One quirk of this approach is that it would make genes for skin pigmentation representational (genes-P), as there has been selection for different skin

pigmentation according to latitude. But it would make most genes for genetic disorders not representational (merely genes-D), because genes for genetic disorders are usually not favoured by natural selection. So at best this would only privilege *some* genes on the basis of representation. We do not have a general rule of representation privileging genes *in general* in any explanations.

Importantly, Shea's analysis applies to any element of the developmental matrix that is selected for its correlation with P. Any non-genetic factors that have a similar developmental role, and have been selected to play that role, also have semantic properties. I emphasize, however, that having a functional role, and being selected to play that functional role is no guarantee of semantic properties. It is important to note that entities can have functions without having semantic content (or to have function without representation as Godfrey-Smith puts it (1999)). Think of the heart, it has a function to pump blood, but has no semantic information content. This sort of discussion does, however, open the door for anything that *does* represent phenotypes (more or less accurately) to take up a privileged role in development. It is possible that some culture, such as a teacher's intention may, indeed, possess this representational quality. Such cultural 'causal representations' will be discussed in chapter seven.

To summarize this section, genes could acquire representational content as a result of selection *for* their correlation with phenotypes. However, this still would not necessarily figure in developmental explanations. Now, I further argue that *even if* there was representational content in the genome that is in some way relevant to developmental processes, then there is no interpreter or consumer of this semantic content and the content, as such there is content, is inert.

### Denial of Genetic Information (3): No Interpreter

No matter how much semantic information or representation there is, it is only relevant to explanation if it is used. Information must be interpreted (Jablonka 2002), consumed (Millikan 1984) or received (Maynard Smith 2000). If there is no such

interpreter of the genome then no amount of genetic information is going to warrant developmental privilege on the grounds of representation of phenotypes. It is the interpreter's function to use the information.

In fact Eva Jablonka (2002; Jablonka & Lamb 2005b) takes things further. She appears to argue that interpreters are necessary conditions for information to be present at all. There is only information 'in' some entity if an interpreter can react to the entity in a range of ways. Interpreters *construct* information from signals, as needed, when needed. Natural selection acts to hone interpreters in order to create the right sort of information. Information is not an intrinsic property of a thing. There is no information without an interpreter, and there is no interpreter without intention. Jablonka explains, 'Only living systems make a source into an informational input. The more complex the organism, the more information it constructs' (Jablonka & Lamb 2005b, p. 588). Information supervenes on some physical systems when an interpreter interprets them in a particular way. Information is epiphenomenal to physical causation, but may form part of explanatory models for interpreting agents. However, we don't need to take things all the way to the controversial position that there is no information without interpreting agents. It will be sufficient to argue that there is no such interpreter of the genome.

Shea (2006) puts it like this, 'to be a consumer system a mechanism must have the function of responding to a range of intermediates with a variety of different outputs' (p. 325). The process of interpreting the genome is a process that requires producers, consumers (interpreters) and actions, users of information, whose success is contingent on correspondence between states of affairs and the signals being interpreted.

Think of another informational system, a set of traffic lights. There are many methods for stopping traffic, for example, a big pile of rocks in the road will stop traffic. In this case it is the causal properties of the rocks, which brings about the outcome. However, when the red light is showing, it is not the brute causal properties of the photons falling on a driver's retina that stop the car, it is the information, the

imperative ‘stop’. There is a selective history at work here in a sense. Because lights at intersections are chosen for their redness, in a sense they become lights-P. But there is also an instruction, ‘stop’, which goes beyond the physical properties of the light. The driver is an interpreting system that can respond to the information provided by the lights with a range of actions. Without the driver to interpret the informational properties of the light, the information is impotent.

Rosenberg likens genes to the ink marks on a piece of paper, which spell ‘cat’. The ink has intentional content about cats only in virtue of human minds attributing that content. The ink marks have intentionality that is *derived* from human minds. The marks get their intentionality from our beliefs. This seems compatible with Jablonka’s approach to information, which claims that information exists only owing to an interpreting system. So what is supposedly the interpreter of the genome? It cannot merely be conscious humans, because surely any informational content of genes is not dependent on there being actual humans in existence to interpret them.

The likely candidate for ‘interpreter’ of the genome is the transfer-RNA molecule that binds to the triplet codon of the DNA and brings the appropriate amino acid molecule into place in the growing protein chain (see figure 7). tRNA doesn’t ‘react’ to the genome with a range of actions, it merely bumps into it. It can’t do anything else. If there are no different possible actions on the part of tRNA then it is not reacting to a signal. There is as much meaning, from the tRNA’s, or even the whole cell’s perspective, in the genome, as there is meaning in a computer program from the computer’s point of view. We can imagine that there is no genetic ‘code’, no information in biology at all, and the physical reaction between tRNA and mRNA will be completely unchanged. This is not the case if we imagine there is no such thing as a writing system and a human agent is presented with written glyphs on a page. In this case the reaction will be utterly changed. This is a crucial asymmetry.

Let’s look closely at what is going on when mRNA allegedly codes for the next amino acid to be added to a growing protein chain. First of all, mRNA does not actively seek the ribosomes it ultimately binds with. At the scale of biologically active molecules

movement is so fast and so chaotic as to be all but random. mRNA when it accidentally happens to pass into the vicinity of a ribosome will bind. This is a chemical fact about the molecules concerned, just as when rain drops happen to fall on salt crystals the crystals dissolve. The binding is predictable of course just because of the sheer statistical probability of these two rapidly moving molecules colliding in the confined space of the cell.

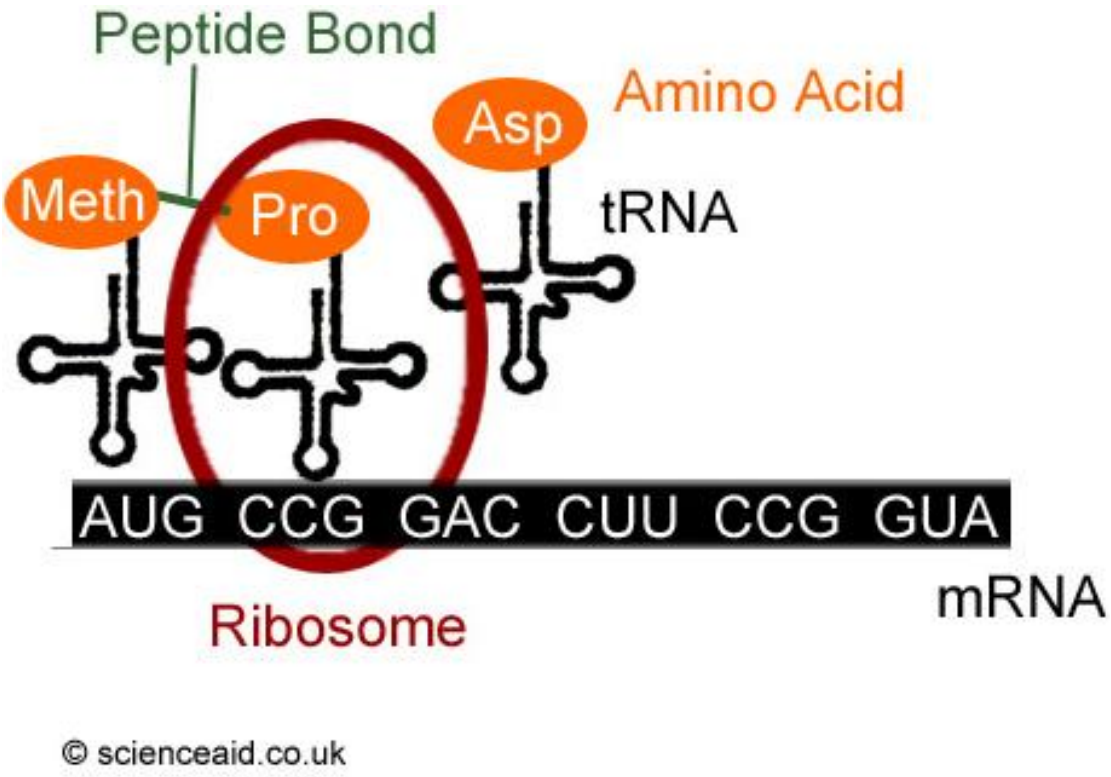



Figure 7. The Relationship Between mRNA and the Protein Chain: As the ribosome moves along the mRNA, transfer-RNAs (depicted as black crosses) bind to triplets of mRNA. This aligns amino acids and permits peptide bonds to form between the amino acids of a growing protein chain.

Next the ribosome moves along the mRNA, much like a drop of water moves down an icicle predictably towards the tip due to the structure of the molecules and solids involved. tRNA binds to the ribosome and, just like the soles of particular shoes tend to accumulate some sorts of pebbles or leaves, particular tRNAs carry particular sorts of amino acids. Only the tRNA carrying the 'right' amino acid can physically bind to the matching mRNA codon. Particular physical characteristics of the interaction between tRNA, ribosomes and mRNA ensure that the tRNA with the anticodon to the

mRNA codon almost always align. Think of another physical system, that operating when a snooker ball drops into a pocket. The mechanism is such that if the ball is white then due to its internal magnet (or sometimes size difference) then the mechanism delivers the white ball to one end of the table and other balls to the other end. Both these mechanisms operate perfectly well even if there were no information anywhere in the systems concerned.

Unlike the case of the traffic light where it is the informational properties of the light that caused the traffic to stop, in the case of tRNA is the physico-chemical properties of the nucleotides binding and the juxtaposition of two amino acids, followed by the formation of a peptide bond between them, according to physico-chemical laws. There is no causal role for information in the genome here, merely a physical role. tRNA does not interpret the genome, it only has one possible action. Someone might argue that what is important in this interaction is the arbitrariness of the tRNA to amino acid relation. Any tRNA (and hence any mRNA codon) could in principle have been matched with any amino acid. However, it actually appears that there are good physico-chemical reasons why some codons map to some amino acids rather than others (Bergstrom & Rosvall 2009) and hence the mapping is not as arbitrary as once thought. Furthermore, I accept that there is definitely a systematic mapping from codon to amino acid, which can only be called a code. But this code exists only between mRNA and the primary structure of proteins. Any developmental privilege warranted by the genome on this basis does not even 'reach' as far as the three-dimensional folding of the protein chain, which is context-dependent and not 'coded'. Finally, Rosenberg's argument against codons representing amino acids (because they do not satisfy the usual criteria for intentionality) suggests that this genetic 'code' is not one bearing semantic information. It is the causal properties of DNA that are important in development, not any informational properties. With the cell there are individual causal components that explain all the workings and development. Even if we could see the whole system as 'using' the information *we don't need to posit information to do so*.

We must understand an important point made by Godfrey-Smith (2000, 2006a). In its original context the theoretical role of genetic coding was restricted to the idea that mRNA specifies the amino acid sequence of *de novo* proteins to be folded in the cell. That was it and that was all. Even the 3-dimensional secondary, tertiary and quaternary structures of proteins are dependent on local conditions outside of genetic control. And the subsequent involvement of proteins in cellular biochemical reactions is contingent of a host of other particulars, to say nothing of high-level phenotypes of organisms. Any information in genes could 'reach' only as far as the primary amino acid sequence of proteins. The construction of this sequence is mediated by transfer-RNA molecules (tRNAs). If there is information in the genome it is restricted to the gene-mRNA-tRNA-amino acid chain system. Godfrey-Smith draws this crucial point out by explaining that when you order the extra-large pizza you are not ordering the delivery to be late, even if this is a likely consequence of placing that particular order. The likely (or necessary) effects of a signal are not always part of its content. So it is with genes and development.

What is interpreting the genome? Is it the tRNA? Is it the whole cell? Is it the organism? None of these seems likely, or necessary to explain the action of genes in development. All that exists are predictable regularities of physical causation. As Kitcher (2001) points out, we need not worry about the language of coding (or information) in genetics, nothing would change in biological theory itself.

I'll draw attention to one last theme pushing through all this talk of information, representation and codes. It seems that no matter what the outcome of the analysis with respect to DNA, then any attempt to privilege DNA in developmental explanation is susceptible to the parity argument. There could in principle be other elements of the developmental matrix that bear the same qualities of information, representation, coding or selective history. This may be particularly so when looking to some elements of culture as I will show in coming chapters. It is parity of reasoning that ought to be our starting point for developmental explanation, not information content.

## Summary

Genes definitely possess information according to the mathematical theory, they are also natural signs. But everything in the world, and certainly everything in the developmental matrix can be described in Shannon or natural sign terms. None of this makes genes special. There may be a transmission sense of information in the genome, but again this phylogenetic information does no explanatory work when we are talking of the ontogeny of traits. Attempts to insert semantic information into the genome are hazy at best and there are arguments suggesting that even if this were the case then the information is not consumed in development. DNA may play a coding role, but this appears to be relevant only insofar as we are interested in the codon-amino acid link. There is not a well worked out informational concept of the genome that shows why and how it plays a distinctive, privileged role in development. It may be that there is a defensible, informational concept of the special role of genes in development, but there is every reason to pursue an alternative approach to assigning importance to developmental factors.

Perhaps this is all just a terminological dispute over information. Well if that's all that is going on here, then this is even more reason to drop the talk of information in genes and to apply the parity of reasoning principle. In the final analysis it is the causal properties of genes that are important not the semantic ones. Hopefully I have shown that it is questionable whether there are any semantic properties. Just like an ant trail doesn't represent anything without an interpreting agent. Just like a pool ball will come out one end if it's white, or the other if it's coloured. The ball doesn't have any semantic properties, it just is a part in a complex mechanism which has reliable causes and effects.

By stripping away the notions of information and representation, then the relationship between genes and development is characterized predominantly by a *causality* relation, which has biological effects, which potentially transfer to the next generation, which are based on a clear mechanism and not on information content. The informational attack on the parity thesis fails for now.

Yet, empirically, DNA is an important cause in development. This is demonstrated by the genetic basis of many diseases, and sex determination in humans. The remainder of chapter four shows how we can retain the developmental privilege of some genes or gene networks with respect to some traits and yet still leave open the possibility of parity with other elements of the developmental matrix. This is because we can take a causal-role analysis of development. This position will lie midway between the genes-are-all and the system-is-everything positions introduced to this point.

## Causal Analysis

Let's return to the project of building a tool for causal analysis that respects the parity of reasoning thesis. A significant criticism of developmental systems theory is that it fails to provide practicing scientists with anything they can actually use. The insistence on taking every causal factor equally seriously is crippling to practical research and especially an obstacle to deciding where and how to intervene on a system. There is a real sense that if everything is important then nothing is. This is the main criticism of developmental systems theory. When deciding how to intervene on a system we generally ask what are the causes and effects here? How do we identify and characterise causes in the developmental system?

Sometimes we are looking to explain a causal chain of events where A causes B and then B causes C and so on. For example, when explaining how a deer died, we might say that first the hunter pulled the trigger, this caused the flint to flare, which caused the gunpowder to ignite, which caused the bullet to fire, which caused the deer to die. However, explanation by causal chain breaks down when more than two variables are involved in cause and effect at any given stage of the process. For example, Rasputin was poisoned, shot, and thrown in the river. How do we explain Rasputin's death? What caused Rasputin to die? It may have been a single causal factor, or two or more in combination. There is not a simple causal chain in situations like these. Furthermore, sometimes there are multiple causes all of which are necessary for an effect. It may be that Rasputin died because he was thrown in the

river, and that his poisoning and shooting had nothing to do with his death. However, in the case of the development of IQ in humans, we have already seen that genes, membranes, and environmental factors all have some role to play.

We need to first, identify the full set of factors responsible for an effect, and then make judgments of causal importance in contexts of multiple, interacting, necessary conditions. Not all necessary factors are equally important.

Often what we really want to know is the structure of a web of relationships. This is particularly true if we are planning interventions on 'critical mediating variables' (Russo 2009, p. 157). Critical mediating variables, or privileged causes, are the sort of thing that policy makers and health professionals are interested in. In psychological development it is not merely the case that A causes B. Or even that A causes B via a set of intermediaries c,d,e… But A and C and D all may cause B due to the structure of their web of relationships. This is the interactionist consensus and the basis for the developmental systems position. We ought to distinguish A from C from D if we can. They may play *different kinds of causal role*. If it can be done, this is a way of structuring developmental systems. So the first thing our analysis of causal roles will need to do is to handle *multiple causal factors*. Which factors do we want to include? There are two options:

1. We could include everything (systemics, developmental systems theory).
2. We could include some things (perhaps a more cognitively tractable approach).

Potochnik (2007) does not think that maximal inclusion is best. Maximal inclusion of causal factors makes for highly complex explanations. These are cognitively demanding and offer little direction for intervention. However, in the interests of making sure we don't leave out any important causes, I will begin with maximal inclusion. The method of causal analysis I propose will then lead us to a tractable set of causes that comprise an explanation. I will now demonstrate how causes can be catalogued and identified as privileged or not. Once we have identified these causes we can build our causal explanatory models. In chapter six I will build one such

explanatory model. Then we can refine the models in ongoing hypothetico-deductive fashion. But the first thing we need to do is to decide what counts as a cause. When is X a cause of Y?

## Woodward and the Minimal Sense of Cause

I will now outline a minimal sense of something being a cause, and then try to elucidate differences between the various factors that might satisfy this minimal criterion with respect to some effect. Once we have a tool for differentiating causes we can then highlight these differences and see how we ought to intervene in order to produce the effects that we desire.

Woodward (2003) proposes an 'interventionist' theory of causation. This is a version of difference-maker theories of causation under which causes make a difference to effects. If we take causal relata to be variables, causation relates changes in one variable to changes in another. In its simplest form the interventionist intuition is that a change in the value of X due to some intervention, would then lead the value of Y to change too. Or, it is at least the case that there is some possible intervention on X that would result in Y changing as well. This is necessary in order that X be a cause of Y. It is crucial that the changes in X and Y occur because of a hypothetical (or actual) intervention on the alleged cause. This is because some relata may be correlated rather than causally related. The classic example of this is the barometer and the storm. Every time the barometer drops there is a storm. But the barometer does not cause the storm, there being a common cause of both effects, which is atmospheric pressure. If we were to intervene (actually or hypothetically) on the barometer then we would observe that the correlations ceased. So the barometer, by the interventionist account, is not a cause of storms. It is a caveat of this theory of causes that the intervention on X is idealized such that any change in Y that results comes about only through the change in X. All other variables that were previously causally relevant to X no longer hold influence over X.

We can consider as a cause any thing C, which if it were to be manipulated in the right way, would produce a change in some effect E. The link between causation and

manipulation (in such cases as there is a link) can tell us something about the causation in question. This approach has been criticized as being anthropocentric because it seems at first pass to limit the domain of investigation to just those causes, which it is possible for humans to manipulate. However, if we characterize our notion of intervention in the right way, usually as some sort of idealized intervention, then we overcome this problem.

Woodward and Hitchcock (2003) explore a way of characterizing the notion of an intervention which does not make reference to the relationship between the variable intervened on and its effects. For Woodward and Hitchcock, in contrast to Pearl (2000), an intervention *I* on a variable *X* with respect to some effect Y, must meet four requirements (M1)–(M4):

> (M1) I must be the only cause of X; i.e., as with Pearl, the intervention must completely disrupt the causal relationship between X and its previous causes so that the value of X is set entirely by I,
>
> (M2) I must not directly cause Y via a route that does not go through X. This can happen in various formulations of a placebo effect example,
>
> (M3) I should not itself be caused by any cause that affects Y via a route that does not go through X, and
>
> (M4) I leaves the values taken by any causes of Y except those that are on the direct path from I to X to Y (should this exist) unchanged.

Woodward then gives a minimal criterion (M) for something being a cause.

> (M) X causes Y iff there are background circumstances B such that if some (single) intervention that changes the value of X (and no other variable) were to occur in B, then Y or the probability distribution of Y would change.
>
> Woodward (2010)

This minimal formulation of what it is to be a cause is very broad. If anything like the neurodevelopmental schooling hypothesis, or the environmental complexity theory of IQ scores is correct then (M) includes: television, billboards, the internet, slide-rules, pens and paper, membranes, genes, maternal diet, and so on. (M) rules out dolphins as causes of human eye colour, but (M) includes oxygen as the cause of

house fires, the road as a cause of the chicken crossing, and genes, as causes of reading. Such causes are not usually what we are referring to when we ask, 'why did the house catch fire?' or 'how did Sally come to be able to read?' Woodward is looking for some way to take this minimal conception of causation and refine it so that the concept of cause does the explanatory work that we want it to.

Note that this minimal criterion rules out mere correlations. There is no intervention on a barometer that would have any effect on a storm. (M) says nothing about whether a cause tends to produce its effect or inhibit it, and so a factor that tended to diminish or negate some effect would count as a cause of that effect under (M). It is the breadth of this kind of approach that makes it useful. I will begin with this criterion for identifying multiple causal factors of some effect. This ought to satisfy systems theorists, as it will capture everything that is causally potent.

Woodward's next move is to produce three dimensions of causal powers. I will examine in some detail his three 'causal dimensions' and then suggest that we need a fourth if we are to capture (and differentiate) all causes in one model. A four-dimensional 'concept-space' captures all causes we might be interested in when analyzing developmental causation. Finally, I explain how this tool can then identify the privileged causes that we want to build into our explanatory models.

## Kinds of Causal Role

There are different kinds of causes. To supplement the minimal criterion (M) for something being a cause, we can introduce three causal dimensions, stability, proportionality and specificity. We can distinguish the particular role that some causes have with respect to some effects and contrast this with the role that other causes have.

### Stability

The key requirement for causal stability is that the cause in question produces consistent effects over a range of manipulations in background conditions. Stability comes in degrees and causes can therefore be more or less stable than other causes.

For example, a gene that consists of CAG repeats in the Huntington region of chromosome four causes Huntington's disease[9], which is a degenerative neurological disorder characterized by cognitive decline and jerky limb movements (chorea), across a wide range of other combinations of genes, developmental environments and ecological conditions. In fact the huntingtin CAG repeat causes Huntington's disease across a range of background conditions virtually as wide as will permit a cognizing human to develop (Paulsen et al. 2008). The huntingtin CAG repeat is an extremely stable cause of Huntington's disease. It is stable across almost all backgrounds that are relevant to the domain of interest, that being the development of a cognizing human.

On the other hand a mutant version of the KRAS gene causes certain sorts of leukaemia, but only given a host of other important background conditions, such as exposure to a carcinogen, the amount of p53 protein present, and so forth (Zhao et al. 2010). We might say that CAG repeats are highly stable causes of Huntington's disease, whereas KRAS is a less stable cause of leukaemia. The important point is that although CAG repeats and KRAS are both causes of serious disease, they are different kinds of causes because they can be plotted at different locations along a continuum of stability. With, zero representing completely unstable causes and one representing utterly infallible causes (figure 8).

[9] For more information about Huntington's disease and the genes involved see chapter five.

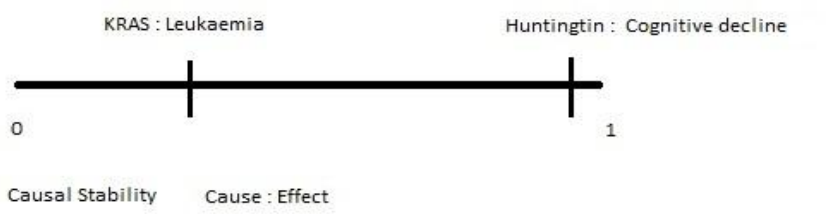



Figure 8: Causal Stability (Cause : Effect)

Another important factor with respect to the stability of a cause is consideration of the proximality of the cause with its effect. It tends to be (though is not always) the case that the more proximal a cause is to its effects (given some chain of causation or a sequence of events) the more stable the cause is likely to be. So if we describe some system and hold the grain of description fixed, then compare two causes, the more proximate cause will tend to be the more stable cause. I'll give an example to illustrate this point. In his discussion of genetic coding Godfrey-Smith (1999) notes the imprecision of the relationship between genes and phenotypes, when compared to the specificity of the relationship between triplet codons of mRNA and the amino acids added to a growing protein chain in translation. Godfrey-Smith claims that this localized precision is one reason to claim that mRNA codes for proteins, while resisting claims that genes code for phenotypes. Of course the causal chain from genes (via translation, splicing, transcription, protein folding, cellular biochemistry and other aspects of development) to phenotype is a long and tortuous one. However the step from mRNA to primary amino acid sequence is very proximal and tends to be much more stable as there are few possible interventions that ordinarily disrupt the causal chain at this point (while still permitting the cell to actually survive, again, our domain of interest).

### Proportionality

The next characteristic of causal role is proportionality. Woodward explains that causal accounts (explanations) can be too broad or too narrow. For example, to say that learning English was part of the cause of a child becoming literate is too narrow an account, as it implies that learning another language would have had a different

effect. It was not English that caused literacy, but the learning of *a* language with a written form (among other contributing causes). Learning *a* language would be the cause that is proportional to the effect of becoming literate.

Causal proportionality can also be attributed overly loosely on the effect side of a set of causes and effects. In fact looseness in the proportionality of causal talk is a common fault of many who make claims about the importance of particular genes. I will describe this phenomenon in general terms then give a concrete example. We may be told that X is the gene for some trait Y, when actually; X is a cause of some general or more fundamental trait Z, which Y depends upon. However, Y itself is subject to many other more particular causes. What I mean by 'particular' is that even given X, there is no guarantee that Y will develop. Y depends on causes A, B and C against a background of X. I believe that this problem is common when talking about the causes of cognitive phenotypes.

For example, take Williams' syndrome, a disorder of many cognitive functions including reading ability and facial recognition (Karmiloff-Smith 2009). The genetic anomaly is a deletion of about twenty-five genes on chromosome seven. But it isn't accurate to say that the genetic differences cause impairment in reading. In a sense they do, but this is as a consequence of a more global impairment in cortical function. By attributing the effect 'impairment in reading' to the cause 'genetic defect in Williams' syndrome', we are being overly broad in causal attribution. The genetic defect in Williams' syndrome is proportional to the more distal effect, global impairments, not to specific proximal abilities such as reading. The same goes for genes like the FOXP2 'language gene', which produces a regulatory protein that binds downstream to all sorts of target genes. FOXP2 does not have domain specific effects[10]. These confusions arise because of the hierarchical organization of biological systems. I believe that this proportionality problem is common when talking about the causes of cognitive phenotypes. Although many genes may appear to be the genes 'for' certain mental traits, they are merely causes of the background conditions enabling these traits, rather than causes of the traits themselves. The genes lack tight

[10] For more detail about FOXP2 see chapter five.

proportionality with the effects. Getting proportionality right in causal explanations allows us to build explanations that generalize.

In an example given by Woodward himself a pigeon is trained to peck at red symbols. When presented with a scarlet symbol the bird pecks. But the bird did not peck because the symbol was scarlet, it pecked because scarlet is a kind of red. To say that scarlet caused the pecking is to exclude important causal information. By explaining the pecking as caused by the symbol being scarlet is to give an explanation that does not generalize well. By giving the explanation that the bird pecked because the scarlet symbol was a kind of red allows generalization. Similarly, in the above example, saying that the genes caused reading difficulties does not generalize to other situations of cognitive deficit as well as claiming that the genes caused general cognitive impairment.

It helps to think of this as a case of a category doing the causing not a sub-category. So the category 'red' is the cause of the pecking, and any instances of it (perhaps sub-categorical instances such as scarlet) will cause the effect. Scarlet is a sub-category of red. In an important sense red subsumes all that participates in a certain section of the visible spectrum. It is this range of visible light, not any component of it, or cause of it, that is important. The point may be made a little clearer by imagining that it was not red that the bird was trained to peck at, but a non-primary colour such as green. If two lights are shone, blue and yellow, in such a way as they produce a green trigger for the bird, then it is not the blue light nor, the yellow light, nor the lights in combination, that cause the bird to peck, it is the green. Though the blue light and the yellow light played a causal role in the pecking, it was not these that were proportional to the effect.

In an analogous way, liking Mozart is not proportional to a gene for perfect pitch, or any other genes, or gene combinations. One likes Mozart because one has a functioning auditory system and this is combined with the hearing of Mozart in favourable circumstances, perhaps as a child. Here we have a proximal-distal distinction again. The proximal cause was hearing Mozart. The distal cause was having

a functioning auditory system, and genes for perfect pitch. The genes are causes of the neural networks that support audition. The auditory cortex supports the concepts of music and pitch.

Causal proportionality can be mapped as a continuum, just like causal stability. Causes are more or less proportional to their effects. Figure 9 shows how we can plot causes against two continua, stability and proportionality. The genetic defect in Williams' syndrome is a fairly stable but non-proportional cause of reading difficulties.

Looking at examples of proportionality, we see a pattern emerging. It is one of layers, or hierarchies of causes and effects. There is a category and sub-category phenomenon, a proximal and distal phenomenon, and also a levels-of-explanation phenomenon.

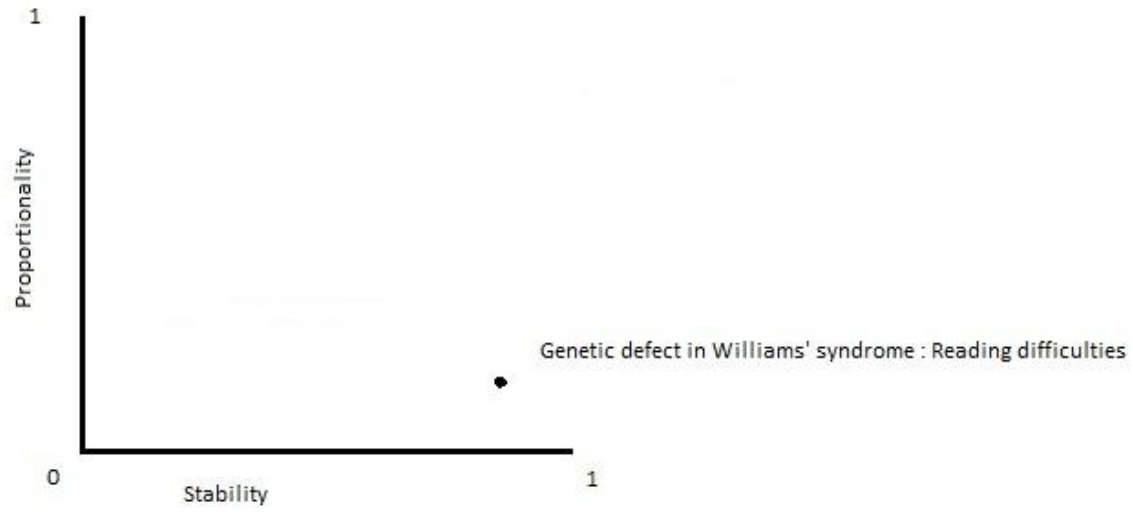



Figure 9: Causal Stability and Proportionality (Causes : Effects)

One of the main potential uses of proportionality distinctions, I think, is to tease apart levels of explanation. For example, high-level sociological (biological/psychological) features like unemployment rate are proportional to high level features such as crime rate. Low-level features like, Jones has no work, and Smith has no work are not

proportional to crime rate. This is partly a value judgment because it tends to result in us favouring causes that are at the 'level of explanatory interest'. Proportionality, therefore, is to some degree a feature of our description of causes and effects. But when our project is one of explanation then we must account for both features in the world and also our interests in those features. If we want to intervene on a system and the cause that we intend to intervene on is not proportional to the effect of interest then we must expect broader or narrower alterations than just those of the effect of interest.

In Huntington's disease there is a situation where a faulty gene (Hg) causes a global cognitive impairment (Gci). The Gci consists of jerky movements (Jm), impaired reasoning (Ir), and memory loss (Ml), among other things. Hg also causes a faulty protein (Fp). Fp is caused by Hg, it is not caused by anything more broad (such as other genes, or a chromosome), or by anything narrower (such as a particular nucleotide base). At the level of biological organization relevant to the synthesis of proteins there is tight proportionality between Hg and Fp. The effect of Hg at this level of description is no broader or narrower than Fp, and the cause of Fp is no broader or narrower than Hg. Take the description to a different grain. Now we are looking at the causes of Gci. Let's limit our cases of interest to only the population of individuals with the huntingtin gene in the first instance. In this population Hg causes Gci. At the level of description of Gci, Hg is a proportional cause because nothing broader or narrower than Hg is responsible. As before, it is not some wider collection of genes or a whole chromosome that is causally responsible, neither is it a single nucleotide base. All of Gci is explained by Hg. At both levels of description, the protein as effect, and Gci as effect, Hg is proportional to the effect. However, take the effect of jerky limb movements. Hg is the cause of Jm. Nothing broader or narrower than Hg causes Jm. So we have proportionality on the cause side. But, we don't have proportionality on the effect side. This is because Hg causes impaired reasoning and memory loss as well. So if our effect of interest is Jm, then Hg is the cause. But if our cause of interest is Hg, then Jm is not the only effect.

Take now the wider population exhibiting cognitive impairment, not just the set of individuals with the huntingtin gene. If we specify this population as the effect of interest then we strike a limitation of causal analysis in general. In order to capture any relationship between global cognitive impairment in general, and its causes, we must necessarily describe the causes in very general terms. For example it is a fact that global cognitive decline is caused by some genetic or environmental factor or other. However, without being more specific about our effect of interest, it is hard to see how we could populate the concept-space. This whole approach requires that we determine a host of important factors ahead of time, such as defining the effect of interest. But once we have determined these factors, and delimited our domain of interest. Then this causal continua approach produces an answer about the relations of causes and effect.

One last point about proportionality ties this concept to the idea of causal mechanisms (something I return to in chapter six). If we want to explain why a Huntington's patient has jerky limb movements, the appropriate fine-grained account would explain the mechanism whereby global cognitive decline interferes with locomotion. If we wanted to explain why the patient had global cognitive decline we would want to know the mechanism by which the huntingtin gene affected neural function. If we want to know why a Williams' syndrome patient can't read, we want to understand that mechanism whereby global cognitive impairment interferes with experiences that would otherwise result in reading acquisition. Explaining the mechanism through which the genetic defect in Williams' syndrome causes global impairment does not explain reading difficulties. Hence, there is an important relationship between proportionality of cause and effect and the locus of the mechanism of interest. Mechanisms of interest connect causes with proportional effects.

The causal property of proportionality is sensitive to how we describe the effects. But, some precision in our description of the effect of interest is required. No approach will have unlimited generality.

### Specificity

The third dimension of causal kinds for Woodward is the specific cause. Specificity is compared to David Lewis' (2000) notion of influence. Some cause C, has influence over an effect E, to the degree that changes in the cause (C1, C2, C3…) tend to produce changes in the effect (E1, E2, E3…). The idea is that a specific cause can lead to fine tuning of the effect (a dial rather than a switch). There is a somewhat precise one-to-one relationship between variants of the cause and variants of the effect.

For example, imagine I walk into a shop and buy a lottery 'scratchy' ticket. The ticket has a range of possible states (win $10,000, win $1000, win $100, win $5, replacement ticket, lose) which cause my wallet to subsequently be fatter or slimmer. In this case the lotto ticket is a specific cause of the state of my wallet at some later time.

On the other hand although the second chromosome in the twenty-third position of a human karyotype causes the sex of the person concerned (pathological cases excepted for illustrative purposes) there are only two alternatives X or Y. So whatever else the chromosomal cause of sex determination in humans is, it is not highly specific in the present sense. There is a sense in which something being more specific allows that cause certain degrees of freedom. The cause can be 'turned up' or 'turned down' in something of an analogue way if it is highly specific. Woodward gives the example of a frontal lobe injury being the cause of someone's low IQ score. The injury is certainly causally responsible, but the effect of the injury can't be fine-tuned up or down in a controllable way. This opens a further question. In order for something to be a specific cause, do the states C1, C2, C3… have to be controllable? I suspect that it is enough that several such states exist (glancing blow, full frontal impact, wounds that crack the skull and those that don't) and that the effects of these causal states show a degree of one-to-one correspondence with a suite of outcomes. This is to be contrasted with causes such as the effect of the speed of light on an optical switch. The speed of light has only one possible state. As such it can never be a specific cause of any time delay in activating an optical switch.

With respect to specificity, a cause and an effect can be related in several ways. There can be many-to-many relations, many-to-one, one-to-many, and one-to-one. In a many-to-many relationship, there are many possible states of the cause and these map onto many possible states of the effect, such as when adjusting a thermostat. Many to one relations hold where many states of a cause map to just a single effect. Many possible charges listed on an arrest warrant for Jones will cause the single effect of Jones being arrested. With one-to-many cause-effect relationships the single state of the cause maps to many different states of the effect. For example, Smith's bachelor degree could cause him to obtain a number of different possible jobs. Finally, with a merely one-to-one mapping there is just one possible state of the cause, and it maps to one possible state of the effect.

A subset of many-to-many functions exists where there are only one-to-one relations between states of the cause and states of the effect. Each state of the cause corresponds to just one state of the effect. This situation of many possible states of the cause mapping to many possible states of the effect, in one-to-one fashion is known as an 'onto' function and must be distinguished from the merely one-to-one relation. In the merely one-to-one mapping there is merely one possible form of the cause and one possible effect. The cause is either present or it is not, and there is only one possible effect, which occurs if the cause is present and fails to occur if the cause is absent. This kind of relationship is a 'switch'. The cause is either there or it is not. Just like a light switch is either on or it is not. But also the sex chromosomes in human beings are close to acting as a switch. If there is a Y chromosome present the phenotype is male, if not then it is female. The Y chromosome is also a proportional cause of maleness.

I am counting switches as low-scoring on the specificity continuum. The Y chromosome is not a very specific cause of maleness, a light switch being 'on', is not a very specific cause of the light being on. These causes are however proportional (and may be stable), so they still have the potential to be privileged causes depending on what questions we are asking and what effects we are interested in. For comparison, the actual genes on the Y chromosome may be specific causes of maleness if there

are variations in genes that map to varying degrees of such things as hirsutism, or baldness, and so on. These genes however may not be proportional to these effects.

Finally, sometimes effects happen and it is alleged that an omission was the cause. For example, the boy died of meningitis because he was not vaccinated. One way of dealing with omissions in this schema is to see an omission as one possible state of a cause. We could take the cause 'vaccination' and it has the possible states of 'being vaccinated at the right time', 'being vaccinated a little late', 'not being vaccinated at all'. In this sense omission would be accounted for under specificity. However, in the case of the boy dying, it seems that he died not because of some omission of vaccination, but rather because he became infected with meningococcal bacteria. If it is possible to explain cause and effect relations without appeal to omissions then this is what I think we should do.

## The Causal Role Concept-Space

We can plot any given developmental cause (or indeed any cause of any effect whatsoever) according to these three continua. The cause can be more or less stable, more or less proportional, and more of less specific with respect to some effect that we have nominated as the target of explanatory interest. Whenever we describe something along a number of dimensions we can plot the resulting concept as a spatial representation. For example we can describe the length of some object along one dimension and plot the result on a number line. We can do the same with something described by two dimensions as is frequently done when describing the relationship between two variables such as height and weight. It is possible to represent concepts with three dimensions as located in a box (see figure 10).

Recently there have been several moves in the philosophical literature away from typological thinking (definitions in terms of necessary and sufficient conditions, or essences). This can be seen, for example, in Godfrey-Smith (2009) and Hobson, Pace-Schott & Stickgold (2000). Rather than summaries of what is necessary or sufficient for something X being a case of Y, we are given a set of descriptive dimensions and X

can be scored as more or less satisfying these dimensions. The idea is that the dimensions create an abstract space in which X can then be located. The position taken up by X in the state-space then indicates to what degree X is a case of Y. Often different regions of the space can be mapped onto different important concepts in some domain. For example the number line might have regions representing 'short' objects and 'long' ones. Though sometimes such concepts will be relative. Godfrey-Smith uses this approach to criticize replicator and 'recipe' conditions for natural selection, and Hobson et al. describes a state-space for something being conscious or not.

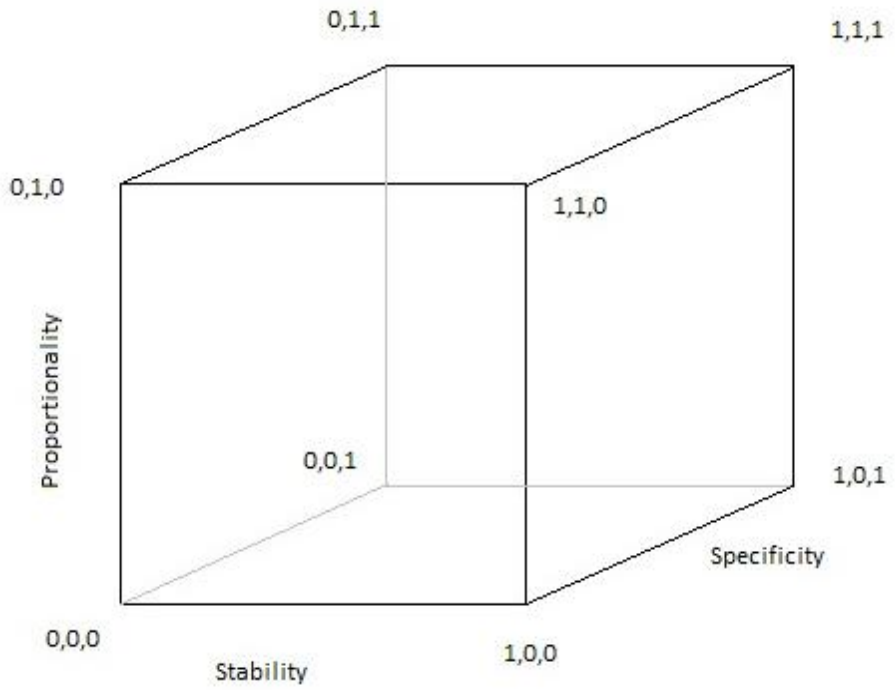



Figure 10: The Concept Space Generated by Woodward's Analysis of Causal Roles

Hobson et al.'s model of consciousness has three descriptive dimensions. These are the level of psychological activation, to what degree the input is external or internal, and a measure of neurotransmitter activity. We can then label different regions of a three dimensional box with concepts such as sleeping, dreaming, waking, unconsciousness, coma, and so on. It is clear that just as we move from waking to sleeping in our lives, a point in the state-space can migrate from one region to

another over time as the measures on the continua change. Different regions of the state-space describe different sorts of consciousness, e.g. dreaming, waking, transitioning to sleep, etc.

Godfrey-Smith's approach to Darwinian populations and natural selection is more complex. At its heart is the idea that regions of multi-dimensional concept-space correspond to populations that are more or less Darwinian. There will be paradigm cases of Darwinian populations, and further removed in space from them, there will be marginal cases, and cases that are not Darwinian populations at all. Instead of the brutality and ambiguity of necessity and sufficiency, an approach usually fraught with outlandish counter-examples, we have a softer 'more or less of a paradigm' approach to something being a case of something else.

The state-space models for consciousness, Godfrey-Smith's 'Darwinian populations', and employing Woodward's causal dimensions as we just have, are all examples of a way of thinking. This is a method that rejects clear distinctions between things in favour of continua. It rejects necessity and sufficiency in favour of degrees of something X being a case of Y. It permits vagueness and marginal cases to legitimately enter our explanatory discussions. Shortly we will employ the concept space for causal roles to identify privileged developmental causes of some trait.

## Models, Abstraction and Explanation

In this section I will briefly introduce models and their role in science. Following this I will identify and solve a problem for the three-dimensional concept space as it stands. I will first distinguish between causal analysis, causal explanations and mechansims, and then discuss one account of what makes a good causal explanation. A causal analysis produces a list of causes of some effect, and ideally distinguishes among them. Woodward's criterion (M) identifies causes of some effect, and the concept-space distinguishes among them. To this point we have been doing causal analysis. An explanatory model is some representation, tool or structure, that assists our understanding of the world or bears some similarity to the way things are

(Odenbaugh 2006), and a mechanism details the entities and the their activities, which produce some phenomenon (Machamer, Darden, & Craver 2000). The present project is one of causal analysis and explanation. I am not setting out to produce mechanisms. The general thrust of my argument is that we can undertake a causal analysis in order to then build an explanatory model, which can then focus our attention upon relevant mechanisms. However, we will see shortly that using just the three causal dimensions described by Woodward leaves both our causal analysis and explanatory model building project wanting. I will describe a way to solve this problem.

Godfrey-Smith (2006b) distinguishes between two processes central to doing science. These are the creation of idealizations and abstractions. With idealization we treat things as having properties that they clearly do not have, such as treating surfaces as frictionless or populations as infinite, in order to study something. With abstraction we leave things out of explanations while still saying something literally true about the world. The project I am undertaking is one of abstraction. Abstraction can both reduce complexity and inject generality into explanations. On one hand the complexity of the developmental system motivates this approach. By leaving things out, we can better understand some complex process. But a search for generality in explanation also motivates this project. Too often explanation under the auspices of developmental systems theory is applicable only to a particular case that we are examining. However, by taking an interest in the specificity and stability of causes, for example, we might be able to generalize explanation in certain ways. With this in mind, we should probably be interested in the different causal dimensions in simpler systems as well. In a later chapter I will connect this hunt for generality and cognitive tractability in explanations with the complementary search for mechanisms through which causes have their effects. My analysis is one focusing on the dependence of variables, not of events causing events, however sometimes we are interested in such event causation and this is when we must appeal to mechanisms.

Abstract models come in various sorts. Models can be *analogies*, such as the billiard ball model of Newtonian physics. They can be relational structures, which purport to

be *isomorphic* with the world, or *similar* in some respects to the world. Often these are diagrams or graphical representations of some kind. Models can also be mediators, in this case they function as a representation that allows someone to learn something about theory and phenomena in a way that is partly independent from both. The third kind of model should not be viewed as a claim about the structure and nature of reality, but rather as a tool in a scientific toolbox (Cartwright, Shomar, & Suarez, 1995). I am predominantly interested in analogies, isomorphisms and similarities. The method I am proposing should lead us to say something true about the systems we are interested in.

Models can help us understand the structure of the world by representing, to a greater or lesser degree of similarity, the nature of things. Models can help us understand causal relations, and deducing causal relations can help us build models. Explaining exactly how the different entities in a model interact to produce some effect is the domain of mechanisms. I will discuss mechanisms in a later section.

Giere (1988) has an account of how model-based science works. A model is an abstract, idealized object, which can be used to represent an empirical system. There is never actual isomorphism with the world, and hence the empirical system never in fact satisfies the conditions of the model. This is a necessary limitation on descriptions of the world, they never correspond exactly in every detail to the way things are. Greater or lesser degrees of similarity are the important relation. At the extreme, we have developmental systems theory, the models of which would include everything in the system in question. At the other extreme we have single factor models that purport to explain, but are very dissimilar to the state of the actual world. So how do we determine relevant similarity, because any two things are similar in some respect? I propose that we analyze the entities in question for causal privilege, and the more of the privileged causes we include in our model, and the more we abstract away from the unprivileged causes, the more relevantly similar our model will be to the system of interest.

A good explanation of some phenomena should clearly explain the phenomena. This means that the set of entities that the explanation posits as being important should include all the important entities. The explanation should not include too many things that are not important, or do not make a lot of difference to the outcome of interest. But if some entity is important or does make a lot of difference then it should be considered and included.

Potochnik (2007) concludes that a good causal explanation of some effect E is one which:

1. Represents the causes of E that figure into the causal relationship of interest in the particular context of inquiry at hand (a limiting criteria).
2. Satisfies the criteria of explanatory adequacy (see below, an inclusive criteria).
3. Is maximally general within these constraints.

She further explains,

> 'The best explanation of an event must, it seems, cite some set of factors that actually does account for the occurrence of the event… if selection would have resulted in the members of the population having trait P1, but because of a lack of genetic variability they instead have trait P2, then information about the selection pressure cannot by itself fully explain the population having trait P2' (Potochnik, 2007, p. 684).

She proposes the following criteria for explanatory adequacy:

4. $Pr(E|C_{expl}) \approx Pr(E|C)$

5. $Pr(E|C_{expl}) \approx Pr(E|C_{expl} \wedge C_k)$ for all $C_k$

In natural language this says that for a good explanation, the probability of the effect E, given the factors proposed in the explanation should be approximately equal to the probability of the effect given all possible explanatory factors (this is criterion 4). And, that the probability of the effect given the causal explanation should be

approximately equal to the probability of the effect given the factors in the causal explanation and any other factor k (this is criterion 5).

If something like this is the right approach to causal explanation then the upshot is that a fully satisfactory explanation should not exclude information that had it been included would drastically change the expected probability of the event to be explained. And we should include only those causes that account for the event and not a host of other minor contributing factors. The three-dimensional concept space is populated with everything satisfying (M), so satisfying Potochnik's criteria (2), (4) and (5) should not be a problem. What we will need eventually is some method for restricting the analysis to only the causes of interest, with maximum explanatory and practical power, i.e. a way to satisfy (1) and (3).

However, this approach still does not distinguish among the causes. In fact when asked for an explanation of why the chicken crossed the road Potochnik would be compelled to include the road as part of the explanation. The road certainly enables the chicken to cross, but we are not always interested in this sort of cause (this is not to say we are never interested in enablers when asking certain questions). Nor are we usually interested in gravity as a cause of bone density, nor background solar radiation as a cause of cancer (except in specific contexts, e.g. astronauts' bones, or airline pilots' cancers.) However, all this falls out with criteria (1) above, 'the particular context of inquiry'. We can limit our causes to the ones of relevant interest later. We want to make sure we actually capture them all in the concept space first. We must also balance all this with cognitive tractability and the pragmatics of what we intend to do with the explanation.

Important, or 'privileged' causes will be found in some regions of the three-dimensional concept-space. Intuitively we see already that a cause that scores highly on stability, specificity and is proportional to its effect is going to be more interesting, important or privileged than a non-stable, non-proportional, non-specific cause (i.e. those that lie at 0,0,0 in the three-dimensional box).

## A Problem for the Three-Dimensional Concept Space

If we are going to analyze and distinguish causal roles in a way that will help to focus our attention on important causal relationships then we want our model to include the important causal entities. If we claim that important causes are those that score highly on Woodward's causal dimensions, then we want intuitively important causes to score on *at least* one dimension. It is demonstrably the case that some important causes fail to score on any of Woodward's causal dimensions. Either this means that we are mistaken in thinking that such causes are important, or the three-dimensional concept space needs modifying. This is a problem, and I will shortly set out a way to solve this problem. There are non-stable, non-proportional, non-specific causes that need to be included in explanatory models. We need to identify these somehow and distinguish such causes from non-stable, non-specific, non-proportional causes that should not be included in explanatory models.

What I will now show is that there is a distinct, explanatorily important, kind of cause that is not captured by Woodward's three dimensions. This is the kind of cause, or causal role, described as 'an enabling cause'. Causes of this kind often cluster around 0,0,0 in the concept space. There are other relatively minor causes that also cluster here and so we need a fourth dimension to pull the interesting enabling causes apart from these. Indeed, we need to pull enabling causes away from 0,0,0 simply in order that they have a chance of featuring in explanatory models. I will argue that we can define to what degree a cause is an enabling cause and map this on a continuum. This continuum becomes the fourth dimension of the causal analysis tool.

Many causes satisfying (M) are located at 0,0,0 in the box. They are non-stable, non-specific, and non-proportional with respect to the effect. They satisfy (M) because there is some conceivable intervention on the cause, which would alter the value of the effect. For example we can conceive of massively altering the mass of the moon, such that its gravitational field has some effect in altering the morphology of some developing organism on Earth. But this seems like an outlandish, unlikely, extreme, or irrelevant intervention given our usual contexts of inquiry. The moon's gravity

deserves to stay at 0,0,0 and not figure in our causal models. However, there will be other causes at 0,0,0, which are *necessary enablers* of effects of interest. We want to reserve a space in the model for such causes, *and* to be able to distinguish them from what might be the truly uninteresting causes. This is the problem with Woodward's analysis as it stands. One key blind spot it has is that there are causes located at 0,0,0 (i.e. non-specific, non-stable, non-proportional causes) which not only are we sometimes interested in, but which also can be distinguished from each other and from uninteresting causes.

Let's think of two different causes of the chicken crossing the road. First the road itself, and second a switch that unlocks the chicken's coop. The existence of some road is not proportional to *chickens* crossing it, neither is it specific, nor a stable cause of chickens crossing. However it is a necessary factor. It enables the chicken to cross. So roads satisfy (M). We can imagine an intervention (removing the road) and the effect, the chicken crossing, is altered. Roads score 0,0,0 in the concept space (or fairly near to 0,0,0) and yet they are necessary causes under (M). Think also of a switch rigged so that under some circumstances the switch is activated and this opens the chicken's coop. The switch is not proportional, nor specific, nor a stable cause of the chicken crossing the road. But again it is a necessary enabler. The switch must be activated in order that the effect could occur. The switch scores 0,0,0 in the concept space as well. It is not stable, as conditions could be such that when the chicken is out there is no incentive for it to even approach the road. It is not specific, as there are not different states of the switch corresponding to different states of the crossing of the road. The switch is not proportional, because it enables many other things not just the crossing of the road. But, the switch is clearly important to explanation (as a distal cause) but is located at 0,0,0 in the concept space I have built so far.

Consider also the ambient air temperature as a cause of the chicken crossing the road. Over a wide range of values the temperature has no effect on this outcome. However, perhaps if the air is seriously cold, maybe sub-zero, then chickens tend to stay in their coops even if these are unlocked. The air temperature, when it is very low has a tendency to inhibit the effect. This satisfies (M), because (M) says nothing

about whether causes tend to cause or prevent their effects. Ideal intervention and alteration of outcome are all that is required. So the air temperature is a non-proportional, non-stable, but perhaps slightly specific cause of the effect. This moves the temperature further from the origin in the concept-space than the enabling switch. This appears to be counter-intuitive. It is much more important to explaining the chicken crossing the road that the coop be unlocked than the air be the right temperature.

I suggest that we need to consider these kinds of causes when we come to build an explanatory model. Perhaps the road is not the sort of cause that we want to include in our explanation of why the chicken crossed the road. It is a part of the legitimately assumed background conditions for the effect occurring. This is part of our context of inquiry (Potochnik's criterion 1). I'm happy for such causes to be considered and then left out of explanation at some later point. However, the switch should probably be considered and included. The point is that we don't want to rule these causes out from explanatory models too prematurely. Just because some cause scores 0,0,0 on Woodward's dimensions, doesn't mean that it is irrelevant to explanation.

Potochnik's criteria include enablers, because removing enablers from explanation drops the probability to zero, or near zero. But her criteria do not single out the privileged causes. Her criteria include too much. Woodward's criteria exclude enablers (or at least place them at 0,0,0). If we can describe the degree to which a cause is an enabling cause along a continuum, or at least in binary terms, then this is unsatisfactory.

If enabling is occurring then there is a degree of bottlenecking in the relationship between some causes and some effect. Enablers will often be distal causes, which bring about some effect by being more directly causally responsible for some intermediate effect. And most, or all, causal trajectories to the effect must pass through this enabling cause. This is clearly the case in the chicken and the road. The switch is a distal cause of some effect, the chicken getting out, which is necessary for

the effect of interest to occur. Furthermore, without the switch being activated, the effect cannot occur.

Ideally we would be able to capture and distinguish all causal kinds in one model. I propose that we can do this by adding a fourth causal dimension and extending the model to a four-dimensional space.

## Dependence: The 4th Dimension

Lombard (1992) argues that there are some cases of counter-factual dependence that are not cases of causation. He gives the example of an event causing a match to become dry, which is not supposed to be a cause of a fire being lit. Similarly, there is debate over whether it is the lighting of the stove, or putting the kettle on it, which is the cause of the water boiling. It should be clear now that there is nothing stopping us from saying that the match becoming dry is one kind of cause, and the lighting of the gas another kind of cause (perhaps a more proportional cause of the kettle boiling than the drying of the match). Furthermore, some of these causes are more or less important than others. Some of these causes correspond to different conceptual spaces than others. It seems that rather than argue over what is or isn't a cause, we should place everything that satisfies the minimal criteria (M) into the analysis and then inspect the concept space to identify kinds of causal role.

Lombard further argues that there is, 'a charge of circularity, unless a way can be found to analyze the concept of an enabler without employing the concept of a cause' (p. 319). Actually we need not do this at all. By adding a 'degree of enabling' dimension to our model we can treat enablers as *bona fide* causes, whilst still distinguishing them according to their location in the concept space. Paradigm enablers seem to be (roughly) events, which had they not occurred, would significantly reduce the likelihood of the effect, often to zero, and which also tend not to significantly rate as one or more of the kinds of causes (proportional, specific, stable). This does not mean that causes that are specific, proportional and stable cannot be enablers (often they certainly will be!), but that some causes can be to a large degree pure enablers. This seems to be the case with the chicken coop switch.

Let's look at a case where there is clearly no bottlenecking element. Reading Einstein's 'On the Electrodynamics of Moving Bodies' may cause you to understand the special theory of relativity. But it does not enable this concept, which can be acquired from many other sources. The effect does not *depend* on this cause. There is no bottlenecking, and so no true enabling. On the other hand, think of the relationship between cell membranes and IQ. There is a strong counter-factual dependence of IQ on membranes. Remove membranes and there is no IQ. Membranes are a clear enabling cause of IQ. An intermediate case is that of the relationship between smoking and lung cancer. Almost all lung cancer is caused by smoking (among other factors), but you can still get lung cancer in the absence of cigarettes if you are unlucky. Smoking is largely, but not entirely, an enabler of lung cancer. Woodward identifies enablers as a class of causes, but relegates them to something along the lines of 'background conditions for x'. However, more can be made of this class and it can be turned into a fourth dimension, a continuum.

But what should we call this fourth dimension. These are enabling causes and there is a degree of bottlenecking. We could call the dimension 'degree of bottlenecking', or perhaps 'enablement'. But these are rather clumsy terms. So let's settle for 'dependence'. We will create a four-dimensional concept space with continua of specificity, proportionality, stability, and dependence. Dependence is roughly: Pr(~Y|~X), the probability of the effect not occurring, given the fact that the enabling cause doesn't occur. If you don't smoke, the probability of not getting lung cancer is close to one.

Woodward's criteria exclude pure enablers. Such causes satisfy (M) but then only register at 0,0,0 on the three-dimensional concept space that lacks the dependence dimension. It seems that we want things that satisfy (M) not to cluster at the origin. The whole point is that we are trying to distinguish among them on the basis of importance if we can. It seems strange that the match drying out is an important part of the explanation for the house burning down but it fails to rate on any dimension of cause even though it clearly satisfies (M). It is easy to imagine an intervention on the

match that changes the effect (unlike the outlandish interventions required to alter the effect of the moon's gravity). The whole point of this causal analysis exercise is to distinguish among causes, so if we can tease apart causes clustered at 0,0,0 then we should.

In order to have four distinct and meaningful dimensions in our causal analysis it should be the case that the causal dimensions can vary independently of one another. The degree to which a cause is specific or stable, say, should have little bearing on whether it is an enabler or proportional. Specificity clearly separates from dependence. An effect can depend on some cause without different states of the cause being possible, such as with membranes and IQ. Similarly a cause can be specific without the effect depending on it. An example here is the thickness of my wallet is not dependent on the scratchy ticket, though the ticket is a specific cause of the thickness of my wallet. However, I want to spend a little more time being clear that proportionality and stability are not inextricably tied to dependence.

Someone might argue that enablers are actually found in non-origin regions of the three-dimensional space, which is based just on Woodward's three causal continua. In particular, enablers seem like fairly stable causes that perhaps lack proportionality and specificity. For example, oxygen is a fairly stable cause of houses burning, having its effects across a wide range of background conditions, hence oxygen rates highly on at least one of Woodward's dimensions without having to posit the dependence of fires on oxygen. However, this relationship between stability and dependence does not always occur. For example, high IQ clearly depends on there being cell membranes present, but membranes are not stable causes of high IQ. Many other background factors such as genes, environments and education must be in place.

One might think that there is some relationship between proportionality and enablement. There seems to be an important connection between a cause being enabling and it being non-proportional. Enabling causes seem to be upstream to a capacity that is co-opted for, and essential to, the effect. Hence enablers are often distal causes, which is one frequent feature of non-proportional causes. So, the

objection goes, enablers will tend to not be proportionate. Take the effect of a pile of wood catching fire. A week earlier an event occurred that caused the wood to dry out. The wood drying out is an enabling cause of the pile of wood catching fire. But it is not proportionate. Proportionate causes would be things like, a blowtorch being applied to the pile, or a match being tossed into the pile. The event 'wood drying out' is so distal to the event of interest (the woodpile fire) that it fails proportionality on the general rule (distality/proximality). Therefore, says the objection, the fourth dimension is not independent of proportionality. In fact, it holds an inverse relationship to it. On the other hand Woodward's dimensions can vary independently of each other.

But we can refute this claim by demonstrating enabling proportionality. Here are cases where both dependence *and* proportionality are satisfied. I hope that it is clear that hearing Mozart is a proportionate cause of liking Mozart. One does not like Mozart because one hears music, and one does not like Mozart because one hears the violin, one like Mozart because one hears Mozart. Hearing Mozart is proportional to liking Mozart *and* it is also enabling. So there are counter-examples to the claim that enablers must fail the proportionality criterion. Another cause that is both proportionate and enabling lies in the relationship between codons of DNA and the associated amino acid. This is a proportionate relationship between cause and effect if anything is. It is also an enabling relationship. The amino acid is not added to the growing peptide chain in the absence of the appropriate codon.

I have shown that:

1. Enabling causes are a distinctive, theoretically important class of causes. The chicken coop switch (and other similar cases) demonstrate this. Such causes are important components of the explanations for some events.
2. Such causes often cluster at 0,0,0 in the concept-space when it is mapped out only including the specificity, stability and proportionality dimensions.
3. However, other causes of little importance cluster here too, such as effects that the moon's gravity has over events on Earth.

4. We need a fourth causal dimension in order to pull enabling causes away from 0,0,0 where they would be omitted from causal explanations. We need some measurement by which causes that are purely enablers, and causes that score on the other dimensions as well as being enabling causes, can be measured.
5. Dependence is such a dimension. Effects can depend to greater and lesser degrees on some causes. IQ depends heavily on membranes, whereas understanding the theory of relativity does not depend very much on reading Einstein's original article.

## Using the Model

My full causal analysis tool is a four-dimensional space. However, partial representations of it can be visualised. For example we may plot just the dimensions of stability, dependence and proportionality (there are 4 possible combinations which produce unique three-dimensional models) see figure 11 (A-D).

Let me describe the nature of the three-dimensional box represented in figure 11(D). This is the causal analysis I have introduced with specificity, proportionality and dependence marked on it. There is an origin 0,0,0 in the bottom left hand corner. There are three apices where a cause can score fully on a single dimension. Passing the edge of your hand across the diagram from bottom left apex to top right apex these are the first three apices crossed after the origin. The next three apices crossed are the locations at which a cause could score highly on two dimensions at once. Finally the top right apex represents high degrees of each of the three causal dimensions represented in this three-dimensional version of the model. The richness of the causal kind with respect to an effect increases as we traverse the model from bottom left to top right. The most privileged causes will tend to be found towards the top-right of the model (or whatever the equivalent location is in the four-dimensional space).

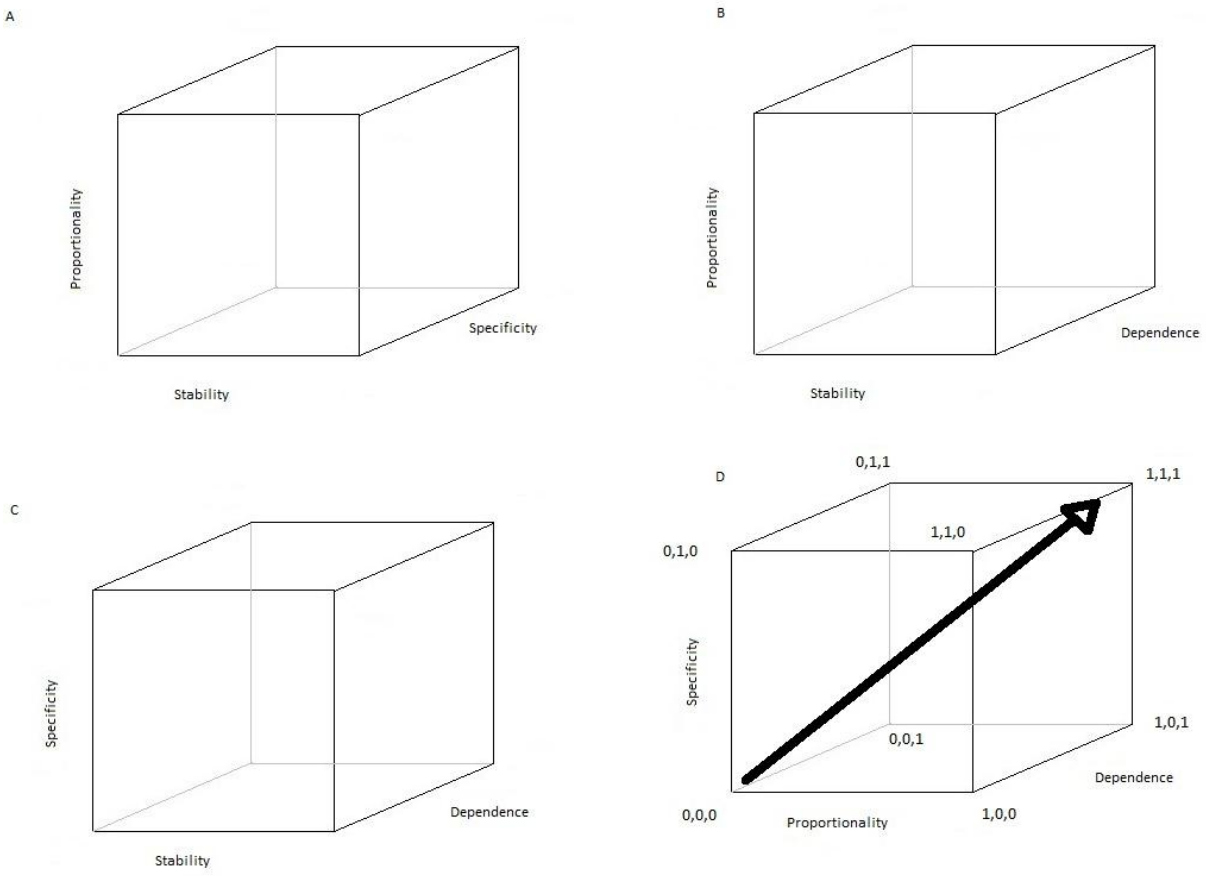



Figure 11: The Four 3D Concept Spaces Showing Partial Representations of the Causal Analysis Tool.

I emphasize two features of the model. First is the fact that the full model is four-dimensional, but there will be equivalent geometric phenomena representing the sorts of spaces I am describing in three dimensions in the four-dimensional version, they are just not as easy to visualise as describe mathematically[11]. The second feature is that it is not obvious how much weight ought to be given to each of the dimensions. Axes could be logarithmic, or perhaps continua are merely binary. Different weightings could be applied to the different dimensions so that movements along degrees of, say, specificity might cause large jumps through the space, whereas movements on other dimensions might not. I am merely taking each dimension of causation as equivalent in worth at this point. There may well exist more and less useful formulations of the concept space. The space could be refined in future work. In introducing this tool I am trying to keep things relatively simple. I am trying to give

[11] For a video rendering of a 4D hypercube see http://vimeo.com/1076061

the outline of a method which illustrates and underscores the importance of cultural technological factors in explaining the development and evolution of the mind, not necessarily a fully refined version on the first pass. Privileged causes are those that tend to be located in the right upper portion of the space when the various three-dimensional representations are visualized.

Another insight I take from Hobson et al. and Godfrey-Smith. In these concept-space models there is not always an obvious transition where some case of causal role X becomes a case of causal role Y. In Godfrey-Smith it is not the case that there is a starting point for 'real' cases of Darwinian populations. Rather the gradient is truly smooth. Natural selection starts playing a more prominent role as we move toward the end of the paradigm. The paradigm of course is a state-space not a linear representation. But still the idea that there is no necessity and sufficiency but rather a spectrum of cases is the key. The same applies to the consciousness model. Just as we slip unnoticeably into and out of sleep and dreams the model allows for no distinct 'phase-boundary' between conscious states. We should expect no distinct boundary between something being or not being a case of natural selection, and no distinct boundary between something being or not being a privileged cause. In fact as we change the question we are asking causes can move around in the space. The FOXP2 gene is one sort of cause of the corresponding amino acid sequence, another sort of cause of language acquisition and yet another sort of cause of lung development.

Taking a concept-space approach to kinds of causal role can help us see now what sort of impact changes to some causes are likely to have on the corresponding effects. Will 'mutations' of cause have radical, subtle, or invisible effects? If a cause is highly specific, then changes to it will likely result in specifically altered effects. If a cause is an enabler, then alterations to it will likely result in cessation of the effects. If a cause is stable, then we can change the background conditions without concern.

Another feature of approachess like this is that we could in theory quantify the position of elements within the space. However, it seems like we are a long way from

actually being able to do this. For now ordinal ordering of causes is sufficient to demonstrate interesting findings.

## Summary of the Argument

1. Either some elements of the developmental matrix are privileged or no elements of the developmental matrix are privileged (all elements could be privileged, but then this is equivalent to none being privileged, so we have a binary disjunction, there are no further options.)

2. The Brute Parity Thesis argues that no elements are developmentally privileged. (Semantic information content could grant developmental privilege, but even genes don't contain semantic information.)

3. Differing causal roles could grant developmental privilege if it turned out that there are appropriately different kinds of causal roles and we can identify asymmetries among causes.

4. Stability, Specificity, Proportionality, and Dependence are appropriately different dimensions of causal roles.

5. To the degree that particular genes satisfy these dimensions of causality, they are developmentally privileged (with respect to the effect in question).

6. But there is no reason to privilege just genes, or even all genes. To the degree that any element of the developmental matrix satisfies the criteria, it is developmentally privileged (this is a Modified Parity Thesis – more in line with Oyama's 'Parity of Reasoning').

7. Therefore, some, but certainly not all, elements of the developmental matrix may be more privileged than others given this gradient. These are the elements in upper right of three-dimensional versions of the state-space boxes.

This approach offers to structure developmental systems theory. It also allows visual inspection of causal kinds, and the possibility for quantification of causal privilege (whatever quantities might mean). Perhaps most interestingly we might now be able to find regions of concept-space that correspond to developmentally interesting concepts, like ‘codes’ or causes that are ‘innate’. This sort of exploration is the material of the next three chapters.

# CHAPTER FIVE – Using the Space: A Focus on Causes

## Introduction

In the last chapter I introduced a four-dimensional concept space of causal roles derived from Woodward's (2010) analysis of biological causation. In order to use this model to build a causal explanation for the purposes of understanding or intervention, we must be able to place causes. Determining where in the four-dimensional 'box' that a particular cause lies is an empirical question, indeed whether it is in the box at all. There are many ways of elucidating the characteristics of a cause. We can use interventional experiments, structural equation modelling, computational modelling, or Bayesian net reasoning to determine whether proposed factors are indeed causes of the effects we are interested in. This will tell us whether (M) is satisfied. We can use such techniques as hypothetico-deductive reasoning and experimentation to refine the description, and place causes in the box.

In this chapter I will detail some of the kinds of genetic and non-genetic causes of cognitive traits and where they might lie in this concept space. Once we populate the concept space with all the causes of some effect of interest, then it will be possible to construct a *practical* model of privileged and relevant causes useful in guiding intervention on some developmental system. This chapter will give examples of the kinds of causes that we actually find in development. I begin with three genetic causes of traits, the huntingtin, FOXP2, and dysbindin genes.

## Genes

### 1. Huntingtin

I have already mentioned the huntingtin gene as a cause of cognitive decline in Huntington's disease. This disease is an autosomal dominant neurodegenerative disorder (Cha & Young 2000). If an individual inherits an affected huntingtin gene then they have virtually a one hundred per cent probability of developing the disease.

The mutant huntingtin gene satisfies (M), the minimal criterion of causation. Furthermore the mutant gene is characterized by a repeated three codon string CAGCAGCAG… The number of CAG repeats corresponds to the severity (or at least age of onset, with longer repeats causing earlier onset) of the disease. Turning 'up or down' the length of the CAG repeat alters the disease in analogue fashion. A second mutated gene can also affect the phenotype and rate of progression of the disease. This gives the gene a *specific* relationship to the disease.

The huntingtin gene is also a *stable* cause, having its effects over a range of background conditions virtually as wide as those producing a functioning brain. Furthermore, the huntingtin gene accounts for virtually all the probability variance between cases where the disease does not manifest and cases where it does, escaped only by dying young of other causes before the disease manifests. Looking to Potochnik's criteria for explanation, we see that the probability of the effect (characteristic cognitive decline) given the explanation that omits the huntingtin gene is virtually zero. So the huntingtin gene *enables* Huntington's disease.

We can describe the 'effect' globally as 'Huntington's Disease', or locally, perhaps just cognitive decline (sans the jerky movements) or just jerky movements (sans cognitive decline), or even just the mutant protein. Two points to note here are first, once we fix the description of the effect we have fixed the causal analysis. And second, the proportionality that the gene holds may change as we re-describe the effect. However, no matter how we describe the effect, the gene is still a stable, specific, enabling, and to some degree proportional cause of the effect. Hence it is a rather paradigm case of privileged biological causation. The gene is the clear target of any interventionist attempt to subvert the course of the disease. I emphasize that genes are very scaffolded things themselves. A lot of causally relevant things go into enabling genes to be translated. For example, membranes, transfer-RNAs, transcription factors, and so on. Transitively these are all causes of Huntington's disease. So huntingtin is certainly not *the* definitive cause of Huntington's disease. However it is the only privileged cause. It is also the factor that explains the variation

between individuals in things such as the presence or absence of the disease and the age of onset.

### 2. FOXP2 – The 'Language Gene'

Locke & Bogin (2006) discuss human life history and language. They describe the KE family in which a significant grammatical and verbal dyspraxia disorder seems to be caused by a mutant FOXP2 gene. There seems to be some *specificity* between states of the FOXP2 gene and states of language use. Karmiloff-Smith (2009) emphasizes that the disorder is wide in its effects (it affects morphogenesis, neurite growth, axon guidance, etc), and that FOXP2 is therefore *not proportional* (in the present terminology) to language function. As further evidence of this Shu et al. (2007) describe the role of FOXP2 in the development of lung and oesophagus tissue. So it seems that if FOXP2 is evolved to contribute to language production then the effect is more generalized and perhaps co-opted from its effects on development of other body systems. Well-entrenched lung development is surely more primitive than language development and this suggests that there was exaptation in the lung to language direction rather than vice versa. Nevertheless, FOXP2 variation does produce variation in language production, it also seems that a normal FOXP2 is required for normal language development (although there is still empirical uncertainty over this *enabling* role because it is not clear what the effects of a deletion of FOXP2 would cause). Furthermore, FOXP2 will have its noticeable effects on language production across a fairly wide range of background conditions because language itself emerges in development across almost all contexts that produce cognizing humans, therefore FOXP2 is a fairly *stable* cause of language function. So FOXP2 is a semi-specific, stable, probably enabling, but non-proportional cause of language production.

### 3. Dysbindin

The dysbindin gene maps to dystrobrevin binding protein 1 (DTNBP1). This protein is expressed in many brain regions and is known to play some role in synaptic

transmission and neural growth. DTNBP1 is characterized as a susceptibility gene for schizophrenia because several mutations of the gene have been associated with the disease (Hashimoto et al. 2009). Various versions of DTNBP1 have also been associated with a range of cognitive abilities in both healthy individuals and schizophrenic patients (Zhang et al. 2010, Zinkstok et al. 2007). It is possible that schizophrenia susceptibility arises partly due to DTNBP1's influence on cognitive performance. For the purposes of my argument I am assuming that these associations are causal, if they turn out not to be then a different example would suffice to illustrate the point. Again I emphasize that determining causation (whether some factor satisfies (M) or not) is an empirical issue.

Certain single nucleotide polymorphisms (instances where one base pairing of DNA is altered) have been demonstrated to both negatively and positively impact on functions such as memory and IQ in otherwise healthy subjects. Burdick et al. (2006) found an association between g and a particular version of the DTNBP1 gene. g is a general intelligence factor. It was discovered by Spearman (1904), who noted a trend for a range of intelligence tests to correlate with each other. Spearman found that a common factor 'g' explained the positive correlations among the tests. g is widely accepted as a measure of general intelligence. Burdick et al.'s study found that DTNBP1 variations accounted for about three per cent of the overall variance in g. Other authors (Zinkstok et al. 2007) have found effects with several other single nucleotide polymorphisms (mutations) of this gene. If there is a mapping from several different states of the gene to several different IQ phenotypes then this would be a case of *specificity* of causation.

Dysbindin accounts for approximately three per cent of the variation in IQ or g, but it is hard to draw conclusions about its role as an enabler without further information. We cannot conclusively claim it is not an enabler because we don't know what the result of having a deleted dysbindin gene would be, neither should we jump from a minimal variation in the effect given variation in the cause, to a claim of no enabling. Dysbindin could still be a switch for IQ (and hence an enabler) even if it is not a dial (it's only partially specific). Furthermore, dysbindin is clearly not *proportional* to IQ as

its effects are shown to involve neurotransmission at neural junctions and broad effects on neural growth. The effects of dysbindin change according to whether the individual is schizophrenic or not (Hashimoto et al. 2009), so dysbindin is *not a stable* cause of IQ or g. Overall, dysbindin is a partially specific, non-proportional, non-stable, perhaps enabling, cause of IQ.

## Placing Causes in the Concept Space

Figure 12 describes the relationship between the genes dysbindin, FOXP2 and huntingtin, and some selected effects. These causes have been plotted in a three-dimensional version of the four-dimensional concept space. We can see immediately, by visual inspection, that huntingtin is a different kind of cause of Huntington's disease than FOXP2 is of language acquisition, and from that which dysbindin is of intelligence.

However, dysbindin is also a gene related to the amino acid sequence of the DTNBP1 protein in a proportional, stable, specific, and enabling way. In fact all three of the genes just described are causes of their amino acid sequences in the same fashion (figure 13). I emphasize at this point that the examples I choose are merely to illustrate the potential of this tool, and I am happy to quibble with respect to empirical evidence, over where exactly on any particular continua any particular cause lies.

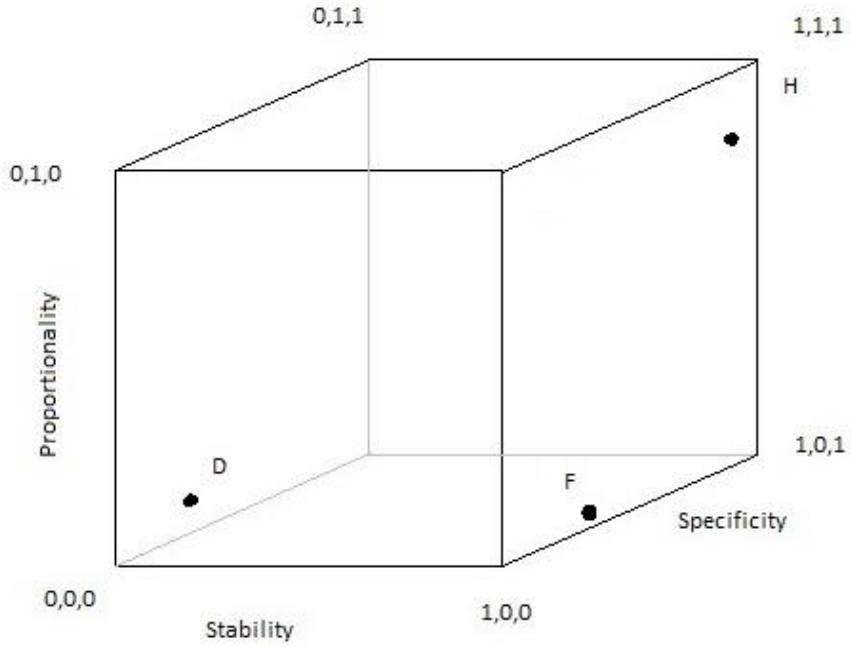



Figure 12: Some Genes as Causes of Some Traits. D represents the causal relationship between dysbindin and intelligence, F between FOXP2 and language, and H between huntingtin and Huntington's Disease.

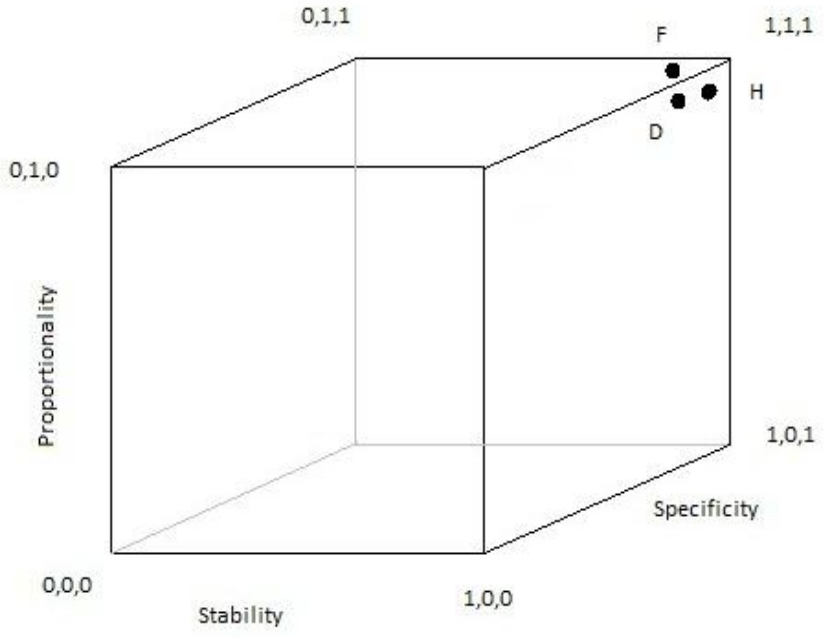



Figure 13: Some Genes as Causes of Amino Acid Sequences. H represents the causal relationship between huntingtin and the huntingtin protein amino acid sequence, F is that between FOXP2 and the FOXP2 protein's amino acid sequence, and D shows the relationship between the dysbindin gene and the amino sequence of dystrobrevin binding protein.

We can now see why the genome is so intuitively privileged in many causal explanations. In many cases, genes are located in the upper right of our concept-space boxes. This is especially true with respect to the causal relationship between genes and amino acid sequences, but it is also true in some 'higher' cases. For example, the huntingtin gene sequence seems to remain proportional to global cognitive decline.

However, this sort of analysis also explains why many genes are not privileged in development. For example, the dysbindin gene is privileged with respect to the amino acid sequence of DTNBP1, but is not in the upper right of the concept space models for traits such as low psychometric test scores (intelligence). This is partly because it is not proportional to this effect. If we ask why someone scored low on a given test, then the answer may be because they have a diffuse brain impairment, and maybe partly because of their particular education history, and perhaps because of a lack of engagement with representational media. The dysbindin gene is not proportional to the cognitive trait, just like the perfect pitch gene is not proportional to liking Mozart in Woodward's example. With the huntingtin CAG repeats, however, the CAG repeats are proportional to the effects at all levels, right up to the cognitive decline. There is a very real sense in which the cognitive decline results from the gene.

Specificity is a phenomenon noted particularly with respect to DNA. Small alterations in nucleic acids tend to produce a variety of small effects on the primary protein sequence. This tends not to be the case with other elements of the causal matrix of proteins that merely satisfy the minimal criteria (M). Small alterations to DNA polymerase or ribosome subunits, for example, tend to render them inoperative. They act more like switches, turning a process on when they are intact and present, but off if they are altered in any way. This causal specificity is one reason perhaps to privilege DNA in causal explanations in biology. Furthermore, in ordinary cases it is the DNA that tends to vary, not the proteins of transcription.

But we privilege DNA as a cause of what? If we consider Godfrey-Smith's arguments for the limited role for coding in biology (chapter four), and Woodward's notions of

proportionality and stability, we see that the appropriateness of causal privilege talk with respect to DNA may have to be limited in important ways.

This analysis explains why (some) genes are causally privileged with respect to (some) traits without positing information content of the genome. What this approach suggests we do is to forget about the semantic content of a gene and decide instead what kind of causal role it is playing. We see, therefore, that some genes are more causally privileged than other genes according to what kinds of effects we are interested in. In other words genes can be causally privileged with respect to some effects and not others. I note that I have not mentioned gene combinations or gene-networks here. But the same analysis can be performed with clusters of causal elements as can be performed with individual elements. It may turn out that some clusters of genes ought to be granted causal privilege, and again this will be with respect to some effects we are interested in and not necessarily with respect to others. On the other hand I will proceed to demonstrate that a lot more things than genes ought to be granted causal privilege according to this understanding of developmental causation.

## What This Shows about the Role of DNA

I emphasize again that the characterization of each causal factor (in this case the three genes I have just described) is open to empirical refinement, however, what does seem to be clear is that the role of DNA in development is pluralistic. There are different kinds of genetic causes. Some are more privileged than others. But there is no universal argument supporting the privileged role of genes *in general* in development (in these cases cognitive development).

Finally, the role of particular genes as causal kinds may place them in the same space in the concept model as other non-genetic causes. For any given effect, and any given region of the concept space, it is possible that both genetic and non-genetic causes might be found there. In the next section I will argue that many other developmental causes play the same kind of causal role as genes in many circumstances.

DNA does not play a distinctive role in development. It plays many roles and even its important roles are equivalent to some other elements of the developmental matrix. Yet, it is not the case that every element of the developmental matrix (nor even every gene) should be seen as on a par.

## Parity of Reasoning

In chapter four, I explained the parity theses of developmental systems theory and argued that we should employ the thesis of parity of reasoning when approaching causal elements. I have just applied the four-dimensional causal analysis to genetic causes of cognitive development and now I will apply the same reasoning to non-genetic causes, including cultural technological causes of cognitive development.

### Number Words

Remember that in chapter one I described the effect number words have on the development of number concepts and, therefore, on the development of exact arithmetic. Number words in a language *enable* exact arithmetic manipulation. Without number words, this simply does not develop (as Piraha and Munduruku speakers demonstrate). Number words have a degree of *specificity*, as there can be none, few or many such terms, and this seems to correlate with different arithmetic phenotypes. Piraha has a 'one, two, many' system, English a set of natural number terms, and A.C. Haddon noted base two counting in Torres Strait tribes in 1889, 'one', 'two', 'two one', 'two two', 'two two one', 'two two two', 'many'.

Number words are *proportional* causes of number concepts. Number concepts are high-level cognitive representations that ride on more basic faculties (we saw this in chapter two). The basic faculties include a low-value exact discrimination system and a higher-quantity approximate system (Feigenson et al. 2004). Many factors conspire to cause these systems, including genes, but once in place, these systems themselves

are enabling causes of the high-value number word concepts. Number words interact directly with these systems to scaffold the development of exactly proportional number concepts. Finally, number words are fairly *stable* causes in this context. We can change the background conditions of individuals around, and yet, so long as they are in a position to learn language, number words will cause number concepts. The existence of the words themselves highlights the fact that there are salient differences between quantities in the world.

Number words are enabling, proportional, fairly stable and fairly specific causes of number concepts and high-value arithmetic manipulations. Number words occupy a location in the concept space for numerical concepts similar to the space occupied by huntingtin genes in the concept space of causes for Huntingtin's disease (figure 14). They are a privileged cause by virtue of scoring highly on several of the causation dimensions. This demonstrates clearly that genes are not the only privileged causes in development. However, have we yet shown that technology deserves causal privilege? Some people may argue that words although causative are not really a cultural technology. I argue that as they are the target of innovation, invention, can exist independently of agents, and can be shared among different users, then they are technological. But now let's look at a more paradigmatic technological cause.

### Storytelling Alice

'Storytelling Alice' is a computer software suite, which is used to teach programming skills. Although people engage with computers on a daily basis, very few learn to create computer programs on their own (Kelleher et al. 2007). Students can use Alice and engage with the software, working through exercises, accepting prompts and clues, and progressively learn to program.

There are two traditional difficulties in getting students to learn programming concepts like loops, parallelisms and conditional logic. These are, firstly, a lack of motivation, and secondly frustrating syntax errors. Alice overcomes these difficulties by being based around the concept of storytelling, and by making it impossible to

produce syntax errors. The storytelling dimension allows users to create animated sequences as the key real-time outcome of the learning sequence. A drag and drop system of commands overcomes the problem of syntax. Users do not need to type the programming commands.

Storytelling Alice enhances student motivation. Users of storytelling Alice were more than three times as likely to sneak extra time to work on their programs. Also, a system of 'stencil' overlays appear on screen to guide users through the tutorial systems while masking functions and commands not relevant to the lesson. The use of stencils greatly decreases the reliance on human assistance and increases the speed that tutorials are completed (Kelleher & Pausch 2005). Students can see the results of their programming and have an active, socially embodied, mode of appreciating their mistakes because they witness the characters in their stories behaving as they expected; or not if they have made an error.

As a cause of skill in computer programming Alice is very powerful. Skill includes problem solving ability, understanding logic, and various other cognitive tools. One study found that at-risk computer-studies students usually average a C and only 47% go on to take the second level computing course[12]. With Alice they average a B and 88% go on to take the advanced course. Taking the advanced course will develop even more cognitive skills. So Alice is a technology that causes (probabilistically in a population of learners) these cognitive skills.

Alice does *not enable* computer skills, since these can easily be attained in the absence of its use. It is a partly *stable* cause because it works for at risk students (i.e. across a variety, but not all, student backgrounds). But Alice is a *specific* cause, as it is able to be modified so as to teach different skills. It is also directly *proportional* to the cognitive outcome. Again, I am certainly not saying that Alice is *the* cause of computer skills and understanding, but it is *a* cause. And we can characterize exactly what *kind* of cause it is (figure 14).

[12] At risk students are those without a math background or prior programming experience.

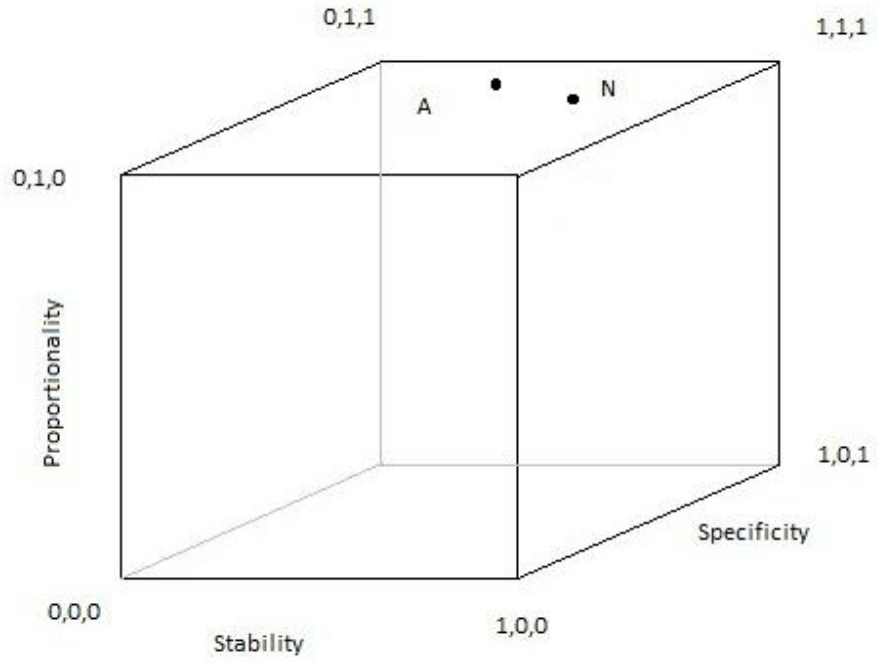



Figure 14: The Causal Relationship between Two Cultural Technologies and Two Cognitive Traits. A represents the relationship between 'Storytelling Alice' and programming skills, N is that between number words and discrete number concepts.

## Summary: Things to note

1. We can take one cause and two different effects of that cause. The cause doesn't necessarily end up in the same place in the concept space with repsect to each effect (e.g. the relationship between a gene and its amino acid sequence when compared to the relationship between the same gene and IQ scores).
2. We can take some location in the concept space and find heterogenous causes located there, for example, genes and cultural technologies (huntingtin and number words).
3. Causes of the same class (such as culture, or genes) are not homogenous kinds (two genes can be in very different locations in the box). Specifying our context of investigation is crucial.

4. We can place *all* causes of some effect in one concept space analysis. How to then select some causes rather than others to build into our explanatory model is the topic of the next chapter.

## Conclusion

In this chapter I have taken the four-dimensional tool for analysing causes, which I introduced in chapter four, and populated it with some representative examples of causes of human psychological traits. By examining examples of genes, and of cultural technologies, I hope that I have made it clear that causes are very heterogeneous kinds. Asymmetries among causes can help us better understand developmental systems. In the next chapter, I take one example of a human cognitive trait, good critical thinking, and undertake a causal analysis, building the concept space for its causes. In principled fashion, I then abstract some causes from the concept space and build a general explanatory model for this trait.

# CHAPTER SIX – Explanatory Models: A Focus on Effects – What explains critical thinking skills?

> 'Causal analysis is concerned with identifying causes and effects of social phenomena [psychological phenomena] with the purpose of understanding, predicting and eventually intervening on society and individuals' (Russo 2010, p. 68).

## Introduction

In chapter four, I introduced an approach to distinguishing causal roles in developmental systems. In chapter five, I showed how some select genetic and non-genetic causes fit into this model. I also pointed out that developmental systems theory is too holistic an approach to be either cognitively tractable or offer up useful tools that practicing scientists and particularly policy makers can actually put to use. As well as the project of explaining cognitive development we are also interested in identifying interventions. Russo (2010) rightly notes that we have two goals when we approach causation in complex systems. One is a cognitive goal, the target is explanation. The other is an action-oriented goal. This is the practical motivation of causal analysis.

Sometimes we want to build simplified causal models in order that we might understand and explain the key determinants of a process or outcome. But we don't want our models so simple that one cause (say a gene) is seen as the only determinant of some effect (say a trait) for this misses other important privileged aspects of the causal relations. In this chapter I will illustrate a method for moving from the four-dimensional analysis of causal roles to a tractable explanatory model and I will illustrate this primarily through a case study of the determinants of critical thinking skills. So our project will need two arms, one where sets of causes are captured, and distinguished, and the other where the relevant or important causes are built into explanations. Chapters four and five began the first project. This chapter introduces the second project.

In cognitive science many of the explanations we seek are layered upon other explanations. For example, explanations of concept acquisition tend to be layered on explanations of brain function, which tend to be layered on explanations of biochemistry and gene expression. These are the issues of proportionality and dependence, which I have already discussed. Depending on the question we are asking different layers of models will be more or less relevant. Relevance and importance is also relative to particular projects of explanation and intervention. I will illustrate this idea of layers or hierarchies in the example to follow.

Furthermore, Oyama (2000) acknowledges that causes in explanatory models can't be too multiple and boundaries can't be too indistinct. We can't extend our explanation to anything and everything in the universe. We need to choose from the very many factors that are causally relevant to the development of psychological traits which ones to build into our explanatory model. I now show how this can be accomplished.

## Critical Thinking Example

By asking what are the causes of good critical thinking skills, or perhaps some more empirically measurable proxy of this, such as scoring well on the California Critical Thinking Skills Test, and employing (M), then we produce a four-dimensional concept space with all the causes accounted for. Once we've fixed a question, then we have fixed cause and effect. To make the four-dimensional causal analysis approach worthwhile, we now need some criteria for taking elements from the concept space and including or excluding them in causal explanatory models. We could merely take any element that we already think is causally privileged and then include, in addition, any other element from the same region of concept space. However, the concept space itself provides guidance as to which causes are more or less privileged and therefore which causes ought to be included in the model. We can include or exclude causes from an explanatory model on principled grounds.

In order to build an explanatory model (with a view to optimal intervention) using the concept-space approach, we first need to list all the causes. As should be obvious

from the preceding chapters, there will be many in the four-dimensional box for any particular developmental outcome. As a toy illustration, let's take just six of the causes of 'critical thinking skills' in turn and examine the kind of cause that each is. We can then populate a causal analysis box (figure 15) with these six causes. I will restrict myself to discussing the three dimensions illustrated in figure 11(D) for simplicity. These are, specificity, proportionality and dependence.

*Dysbindin Genes*: In chapter four, I explained that different polymorphisms of dysbindin relate to lower or higher IQ scores. So if student IQ is a cause of high scores on critical thinking tests then dysbindin shows some specificity toward critical thought. Also, dysbindin may be in some way enabling of critical thought, but this is speculation as it is not clear what the effect of having a deleted dysbindin gene is. Dysbindin is clearly not proportional to particular critical thinking skills however, because its activity is general, at a low level in the biological hierarchy.

*Membranes*: These are essential for critical thought, as without membranes there can be no cognitive function. Critical thought depends on membranes. However, they are not proportional or specific in their effects. In humans, as well as inheriting DNA from both mother and father, an embryo inherits its entire membrane system from the egg. There is a sense in which the membrane system is continuous with the membrane system of the common ancestor of all eukaryotes. This emphasizes one feature of membranes that makes them less significant in explaining the development of critical thinking skills. This is that they are not actual difference makers. Membranes do not actually vary in significant ways in nature (not in the sense that genomes or education syllabi vary in any case), and so explaining critical thinking skills assumes the existence of functioning membranes (this is our context of inquiry, just like assuming the road when asking why the chicken crossed). Membranes are not an explanatory factor relevant to most avenues of inquiry. Membranes as causes are found in a less privileged region of the concept-space model, as they score very poorly on specificity and proportionality. There is neither a many-to-many correspondence of interventions on membranes and various cognitive effects, nor are the effects of the membranes (such as protecting the cell from hazardous

extracellular materials, or facilitating the diffusion of water) proportional to cognitive phenotypes. Most interventions on membranes will cause non-functioning cells, or undetectable effects at the cognitive level.

*High-bandwidth cultural inheritance*[13]: This also enables critical thought, as many of the underpinnings of critical thinking skills are techniques and approaches established in the past by others. These include accumulating symbols, methods and skills that no one person can figure out in their entirety. However, cultural inheritance is less enabling than membranes. A lack of membranes crashes the probability of critical thinking skills to zero. Membranes are a true bottleneck cause. Lack of social inheritance merely makes it very unlikely that some agent will have good critical thinking skills. However, high-bandwidth inheritance is neither specific nor proportional in its effects. This is to be contrasted with the actual skills transmitted, which may be both specific and proportional.

*Literacy*: As I identified in chapter three, the origins, and now existence, of writing may have had a role to play in the emergence of critical thought such as syllogistic reasoning (Lock & Gers 2011). This seems to be supported by experiments, which show that they way students represent a problem maps to how well they solve it (Twardy 2004). Literacy is an enabling, non-proportional cause of critical thought. It is also somewhat specific because the effects depend on the system and method of representation used, for example, Arabic/Roman numerals, representing speech or using pictograms, formal logic, and so on.

*Student IQ*: The specific, and somewhat proportional, effects of student IQ on the emergence of critical thinking skills place it in a privileged region in the analysis of causal roles for critical thinking skills. As this thesis has argued, student IQ itself has an interesting range of biological and non-biological causes, including some cultural technologies (Blair et al. 2005, Flynn 2007). These causes will, by transitivity, be

[13] See chapter seven for a full discussion of this feature of human environments

causes of good outcomes on critical thinking tests, but they will fail the proportionality test (as dysbindin does), and so are likely not to be privileged.

*Critical thinking courses*: Differences in the way critical thinking is taught in the classroom (traditional versus tree-diagramming methods, decision theory, and so on) have effects on how competent students become. The use of computer software for tree-diagramming (van Gelder 2001, Twardy 2004), or even pen and paper exercises (Harrell 2008), have been shown to also mediate this effect. Twardy found that the use of the 'Reason!Able' software suite for argument mapping increased critical thought performance twice as much as a traditional critical thinking course, and eight times as much as standard university courses. A critical thinking course (taken as a whole) is a specific, proportional, but not enabling cause of good critical thinking skills. There are pathways to good critical thinking skills that do not run through the bottleneck of a critical thinking course. Not taking a course does not necessitate a low probability of acquiring critical thinking skills.

Taking just these six causes for illustrative purposes[14], we can now diagram a three-dimensional section of the four-dimensional concept space for causes of good scores on the California Critical Thinking Test (figure 15).

I suggest that there could be good genes, membranes and high-bandwidth inheritance in place and yet the student fails miserably at some test of critical thought. This is because (as this model demonstrates) good genes, membranes and high-bandwidth inheritance are not the privileged causes (toward the top right in this representation[15]) of critical thought. Literacy, IQ, and critical thinking education are

[14] Other causes that will appear in the 4D analysis model might include: natural variation in grey matter, COMT genes, teacher intention, student motivation, all the paraphernalia of apprentice learning, neural prediction systems, pencils, Parental IQ, Informal learning experiences, the questions on the test... etc.

[15] The reader is encouraged to slide the edge of their hand from the bottom left apex of the cube to the top right, as was demonstrated in chapter four, in order to get a feel for the fact that causes scoring highly on none or one of the causal dimensions tend to lie to the bottom left of the figure, and those satisfying two or more tend towards the top right. Again, I emphasize there is no cut-off for when a cause becomes privileged, but some causes are clearly much more or much less privileged than other causes.

such causes. Taking just the causes (from our illustrative selection of six) which seem to be privileged we can build an *explanatory* model of the development of good critical thinking skills (figure 16). Note that I have illustrated the causes IQ and literacy as beneath the effect and the cause critical thinking course as horizontally apposed. This is because there is a sense in which critical thinking skills emerge from the psychologically more basic features IQ and literacy, which in turn ride on features of brains and neural firing, whereas critical thinking courses shape critical thinking skills at a 'higher' level. This is reflected in the more distal, enabling function of IQ and literacy.

We have just moved from an analysis of causal roles (the concept space approach) to an explanation (a box and arrow model). The box and arrow model is an abstraction. It says something literally true about the world while leaving a lot of less important detail out. The box and arrow diagram also generalizes to some degree.

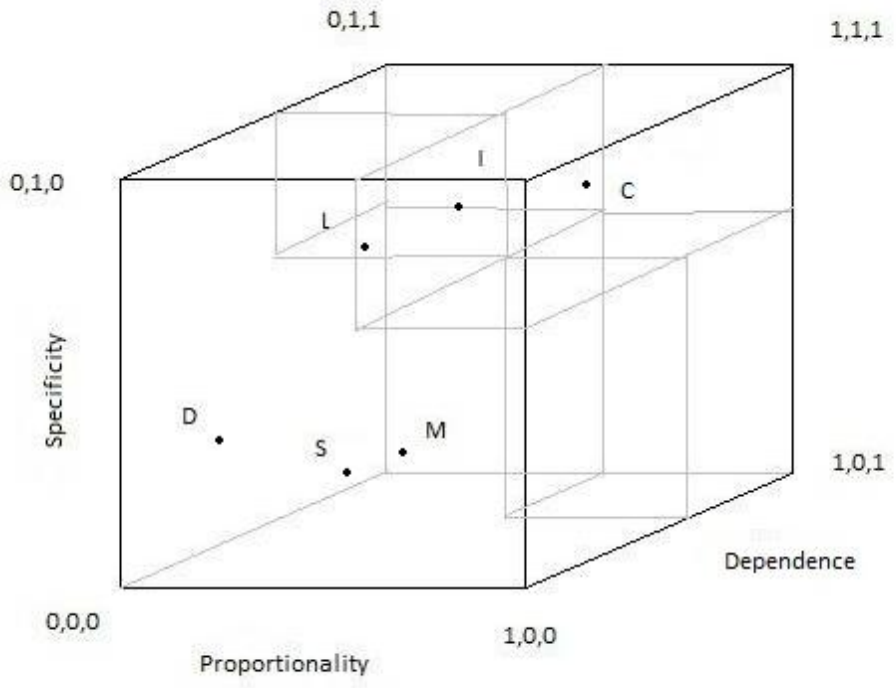



Figure 15: Causes of Critical Thinking Skills. D = dysbindin gene, S = high-bandwidth social learning, M = cell membranes, L = literacy, I = IQ, C = critical thinking course. The light shaded boxes represent the regions of the space where causes score fairly highly on at least two causal dimensions. Such causes are more 'privileged' in this analysis.

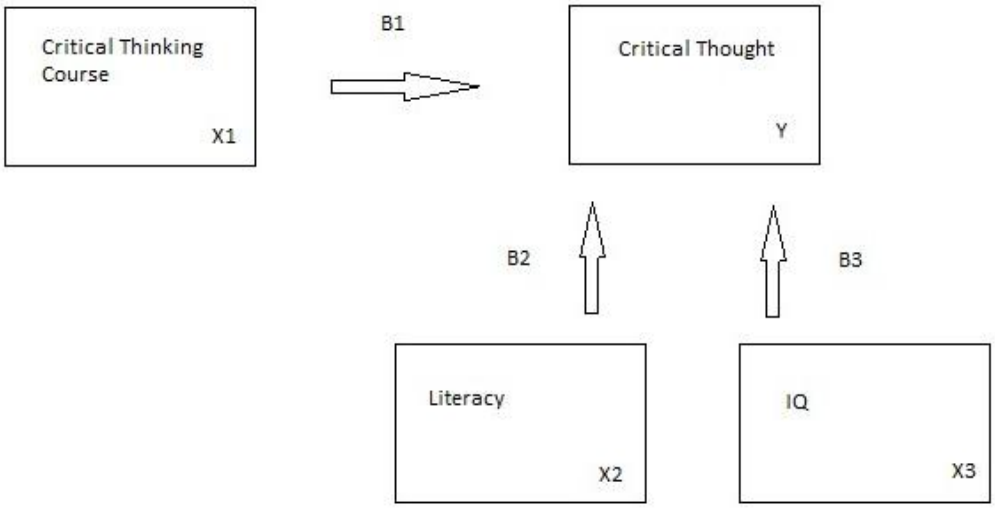



Figure 16: A Causal Model of Critical Thought. This model assumes that the six causes discussed in the text are the only causes of critical thought. The three causes included in the explanatory model are chosen because they fall in the more privileged regions of the causal concept space. The model can be represented as a mathematical model by the equation: $Y = B1.X1 + B2.X2 + B3.X3 + e$

I will now introduce and reply to two potential criticisms of this approach. First, someone might argue that I have *left too much out* of the explanation with the explanatory model in figure 16. Second, someone might argue that figure 16 is not a good explanation because it does not give any *mechanisms* through which the causes identified actually have their effects.

## 1. The Model has not Left Too Much Out

We have left a lot out, but hopefully the process of abstraction has focused our attention upon the causes of interest. Remember Potochnik's criteria for a good causal explanation from chapter four. Part of a good causal explanation was completeness. No element $C_k$ was to be left out, which had it been included would have significantly altered the probability that the effect would occur. The entire four-dimensional concept space analysis satisfies Potochnik's criteria for explanatory adequacy by including everything. But it might be objected that we have failed this test by abstracting to the explanatory box and arrow model. However, as I identified in chapter four, this requires we include the road in the explanation of why the

chicken crossed. So Potochnik's account is too strong. There are things it is reasonable to omit from explanation, either on the basis that such causes are assumed in our context of inquiry, or because such causes are minor players in comparison to other more important causes.

Remember also that models are good or useful by virtue of their similarity relation to the world. An ideal, though perhaps unattainable, model is one that exhibits isomorphism with the way things are. Other models are good because they exhibit relevant similarity with the way things are. This similarity (as Godfrey-Smith explained) can be idealized and abstract. By taking even one privileged cause we generate a model with some similarity to the world. However, by expanding our domain of inclusion from the 1,1,1,1 locus in the causal analysis toward the origin at 0,0,0,0 we can increase or decrease the richness of the model as we see fit. We can build any number of explanatory models with this technique from those that contain only a single causal entity (the *most* privileged cause), to those that encompass the entire set of (M). The richness of the explanatory model we produce is a function of the richness of the model that we consider adequate for our purposes.

In order to sensibly intervene to modify critical thinking as a trait, we need to intervene on the structure of education, on student IQ or on literacy. If we are education policy makers we are mainly interested in critical thinking courses rather than IQ, or perhaps if we are geneticists we are more interested in IQ (and its determinants) than in critical thinking courses. We can omit further causes on pragmatic grounds, as long as we make this explicit.

An explanatory model shows covariate sufficiency when it includes all the variables needed in order to account for the phenomenon of interest. In our toy example I suggest that, to a good explanatory approximation, covariate sufficiency is attained by including student IQ, literacy and critical thinking courses. Long distal chains of causation require only that their most proximate causes be included. If dysbindin has its effect on critical thinking via IQ, then it is sufficient to just include IQ. We aren't

usually interested in the big bang as a cause of psychological traits. However, our domain of interest will vary somewhat depending on the context of inquiry.

It was the complexity and inclusiveness of developmental systems theory, which motivated this approach in the first place. We wanted to simplify the explanation because including all causal elements as equally privileged becomes intractable. I am suggesting that the elements that must *not* be left out are the set of privileged causes. It is these causes that will comprise a *useful* explanation.

## 2. We Don't Always Need Mechanisms

Rather than worry about what is included or excluded in the explanatory model the second objection centres around the relationships among the things that are included. It is true, some might say, that we have produced a model of causal relationships, but this does not *explain* anything because it does not give us mechanisms by which causes have their effects. In effect we have 'black-boxed' the workings of the system. I will give two replies to this objection. First (i), I will show that it is possible to decompose the entities in our model to produce finer grained explanations. Second (ii), I will argue, along with Jackson and Pettit (1992) that it is permissible to screen off detail that is not relevant to the questions at hand.

### i) Decomposition to Mechanisms

Mechanisms are composed of entities and activities. The entities are things in the world, and the activities are what these things do. Mechanisms aim for a complete description of some phenomenon, a set of entities and activities that fully explains the phenomenon without any gaps in the explanation.

> 'Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions' (Machamer, Darden, & Craver 2000).

A mechanism should be able to explain the activities of 'bottom-out' entities *with respect to some field of inquiry*. For example a mechanism in particle physics ought to bottom out at sub-atomic particles. In social science a mechanism should bottom out with the behaviours of individuals, or perhaps their beliefs and desires. We could perhaps explain social science phenomena with respect to particle physics, but I will explain in the next section why we are at liberty not to take our mechanisms that far.

Furthermore, we can distinguish between a mechanism schema and a mechanism sketch. A sketch is more abstract and bottom-out entities do not (yet) feature (Machamer, Darden, & Craver 2000). A sketch can act as a guide to further research and further findings may indicate the sketch is lacking in some important way. It can then be modified. Complete descriptions of mechanisms are continuous without gaps from the start conditions to the termination conditions. The continuity often lies in the arrows in a mechanistic diagram and explication of the arrows is in terms of activities. For example, the mechanism of DNA replication may be represented as a box labelled 'single DNA molecule', an arrow labelled 'replication' and a second box labelled 'duplicate molecules of DNA'. The 'replication' activity is explained in terms of actions. DNA *unwinds*, DNA polymerase protein *moves* along the strand *slotting* new nucleotides into place according to chemical specificities. Note that in the model I have provided there are no such action terms explaining the causal arrows.

However, it is possible to decompose the entities in my explanatory model, such as 'critical thinking course' into parts and activities. The course is composed of teachers, computer exercises and, implicitly, students. Teachers *speak* and encourage, students *write* and *read*, and computer programs *display* correct and incorrect responses. Students *learn*. Once the privileged entities are identified through the four-dimensional causal analysis, then we can focus our attention on the causal arrows of interest and set about deducing mechanisms. Just like the mechanism of DNA replication may make no mention of membranes, or the ambient temperature, both of which might be causally implicated in replication, the sort of causal analysis I have provided allows us to focus on mechanisms without including all that extraneous material. This material distracts from the essential features of the process.

I am identifying causes not mechanisms. It is a model of causes not spatial or temporal relationships. Arrows are not necessarily activities. But building causal explanatory models this way focuses our attention on causal relationships where the mechanisms of interest reside.

### ii) Pluralism of Explanation

When trying to explain phenomena in the social, psychological and biological sciences we can employ several methods. We can choose explanation at different grains or levels of analysis. For example, when asking why crime has risen in some community we could choose to employ the explanation that unemployment has risen. There is a causal connection between the unemployment rate on one hand and the effect, rising crime. However, some might argue that this is too coarse-grained an explanation of this problem. What has actually occurred is that individuals Jones, Smith and Wilson have lost their jobs and as a result stolen money in order to eat. This is a more precise explanation of the effect some might say.

The second explanation is a reductionist explanation. It employs methodological individualism and studies the activities of individuals as explanations of higher-level phenomena. As a general rule reductionists prefer these micro-level accounts rather than macro-level accounts. However, the natural end point of such analysis is that every phenomenon ought to be best explained by reference to particle physics and this is surely wrong. The crime rate has risen because the unemployment rate has risen. Unemployment is importantly proportional to crime. The movements of physical particles certainly take part in these effects, in fact they constitute these effects, but they are not an explanation of the effects, or at least not the sort of explanation we are after in the inquiry at hand. In order to solve the crime problem we aren't out to get Jones a job, and Smith a job, and Wilson a job, we are out to get some people a job and the effect we seek will obtain. In fact had things been different at a smaller grained level, for example it was Roberts rather than Smith who was out

of a job and so stole some money, then the macro-level effect still would have obtained.

Some causes, such as unemployment rates, exhibit realization-insensitivity. This means that it doesn't matter how or why the cause came to be, or how or why a high-level property of some system is physically realized. Given that it *is* realized, the effect will be the same. In the critical thinking case the causal route by which IQ came to be, is not important. It doesn't matter if dysbindin or COMT genes or informal experiences contribute to IQ and in what proportions. What matters is that there is a certain IQ available for the agent to employ in critical reasoning. IQ is the cause the matters, not its particular causal history.

Note also, that (in this case) student IQ is a privileged cause of scoring well in critical thinking tests. But student IQ itself is an effect with myriad developmental causes (we have seen this in chapter one). If we decide to intervene on student IQ we might like to repeat the causal analysis process taking IQ as our effect of interest rather than critical thinking skills, and build a new causal model explaining the privileged causes of IQ. In this way hierarchies of explanation in psychology and development may be built as if one were combining Lego bricks. This seems to be much more cognitively tractable and offers a route to conceptual *understanding* that developmental systems theory fails to strike. This method also avoids simplistic explanation such as genetic or environmental determinist accounts.

I am not arguing that reductionist, or micro-level, or mechanistic explanation is wrong. In fact quite the opposite is true. It *is* the case that causal fundamentalism is true. This means that the activities of fundamental causes result in macro-level effects. If you fix the micro-level then you fix the macro-level. I am arguing that both explanations are correct. Both explanations are useful in different circumstances. Both explanations say something literally true about the world. Pluralism about cause and effect is a legitimate and useful position (Jackson & Pettit 1992). We can reject fine-grain preference in favour of explanatory pluralism and therefore choose which

explanation is appropriate for our needs. This includes choosing between mechanistic explanation and other kinds of explanation.

I want to further argue that we can choose tractable explanations over complete ones. I am not arguing that complete explanations are wrong. They certainly are not. In fact only a complete explanation will be entirely true of any phenomena. But often we are not seeking the complete truth of a matter, but a way of comprehending or a way of guiding our intervention on it. We can acknowledge the force of the developmental systems theory argument without being pushed toward preferring it. We do not throw away an explanation just because there is another. Developmental systems theory is not wrong; it is just not useful for many purposes. Mechanistic explanation is not wrong; it is just that causal analysis allows us to focus on some mechanisms rather than others. Causal models identify which mechanisms we are interested in. Once we understand the privileged causes of some effect we can set about deducing the causal mechanisms that link them.

According to Jackson and Pettit's pluralism we can legitimately screen off lower level detail which is not relevant. In the unemployment and crime case, we do not need to employ the actual mechanism whereby Jones steals bread from the shop, to explain the rise in crime. A higher-level explanation suffices. Including all the mechanisms may give us more realism in our models, but then this limits the generality because the exact same mechanism may not be in operation in relevantly similar cases.

In chapter four I introduced the idea of abstraction. When we build an abstract model we leave things out. When we do this we sometimes make characteristic errors (Godfrey-Smith 2006b). The main characteristic error is that we end up positing some entity that actually doesn't exist in the world. It is a statistic. It might be argued that IQ in my example is merely a statistic. It is not a 'discrete causal player' on its own (to use Godfrey-Smith's terminology, 2006b, p. 7). The actual causal players are the entities that serve as the foundation for the statistic, such as neural firings, enzymes of various efficiency, mental representations, and so on. However, as long as we realize that the entity in question is decomposable to these 'actual players' then we

are not saying anything false about the world. IQ is in the same ontological category as 'unemployment rate' and can serve a similar explanatory purpose.

No initial model will be the final word on causation of any effect we are interested in. We take the causal model and *then* use it for further empirical testing. When we are plotting causes into the four-dimensional model each cause can itself be taken as an effect and its own causes plotted in a further four-dimensional space. Each segment of our causal model can be modified through empirical testing. Refinement comes with iteration of this process.

## Other Examples

Let's compare, for completeness, critical thinking skills to Huntington's disease. See the three-dimensional model for Huntington's (figure 17) and it's causal explanatory model (figure 18). As with critical thinking there are many causes: membranes, the temperature of the womb, and so on… But none of these is anywhere near the privileged zone of the analysis tool. The huntingtin gene is the only such cause. In the Huntington's case we can take explanation directly from gene to trait because the gene is the cause all the way up, at every level, of every manifestation of the disease. Without the gene there would be no mutant protein, no cognitive decline and no jerky limb movements, and each of these effects is sensitive to specificity, enabling and stability with respect to the huntingtin gene. Proportionality holds on the causal side because whether we are talking about the mutated protein that results from the huntingtin gene, or the global cognitive decline of the individual, the cause is the huntingtin gene, no category broader, or narrower. Proportionality may fail on the effect side, if we ask what is the effect of the huntingtin gene then we could substitute many different effects, some broader some narrower. However, I've stated already that we must fix cause and effect in order to build our four-dimensional box, so we can't roam at will up and down through a list of effects. If we were looking for a mechanistic explanation of Huntington's disease then we would need to explore the mechanism by which the huntingtin gene interacts with other elements of the system and the mechanism by which products of these interactions interacted with other

elements of the system. Only then would we have a complete mechanistic account. However, we can give an explanatory account by identifying privileged causation.

Other genetic diseases (or genetically influenced traits) behave differently. Compare phenylketonuria (figure 19), which leaves sufferers severely mentally handicapped. Phenylketonuria is caused by an absence or deficiency of the enzyme phenylalanine hydroxylase. Normally about three-quarters of dietary phenylalanine is converted into tyrosine. If this process cannot occur because of the enzyme deficiency, then phenylalanine accumulates in the body leading to cognitive deficit. In this case both the defective gene and phenylalanine in the diet are needed in the causal model. The gene and phenylalanine in the environment are both proportional to the effect (milk would not be), they are both specific (via different mutations and different amounts in diet respectively). They are both one hundred per cent enabling. They are both somewhat stable because most interventions on the system will leave the effect unchanged. They are not fully stable, though, because we simply change one and the other no longer causes the disease. However, three out of four causal dimensions probably qualifies both the gene and the phenylalanine as privileged causes and warrants their inclusion in a causal model (figures 20).

Finally, I emphasize that the critical thinking skills case is just a toy example for illustrative purposes. I am trying to indicate the manner in which the four-dimensional analysis of causal roles could be used to explain (in a cognitively tractable way) interesting facts about trait development. I am completely open to empirical refinement of the particulars of any one causal model.

## Conclusions

Through chapters one, two and three I built a picture of a very complex set of interacting causes in psychological development. This included technological influences, constructive neural processes, and external resources. The picture was complex even though I was trying to focus only on technological influences on cognitive traits. Explaining the development of human psychological traits is a messy

business. In this chapter I have furthered my argument for bringing tractability to Developmental Systems Theory by using the concept-space method, which I built in chapters four and five, to identify privileged causes of traits and build explanatory models.

I have illustrated the method using the example of the development of critical thinking skills. I have suggested that culture and some cultural technologies play key roles in explaining this aspect of human psychology. In the next chapter I will discuss two theoretical implications of the concept-space tool and then, in the final chapter, I turn from the present to the past, and indicate how understanding the privileged role that human technologies sometimes play in the development of psychological traits, helps us to understand two problems in prehistory and the evolution of human minds.

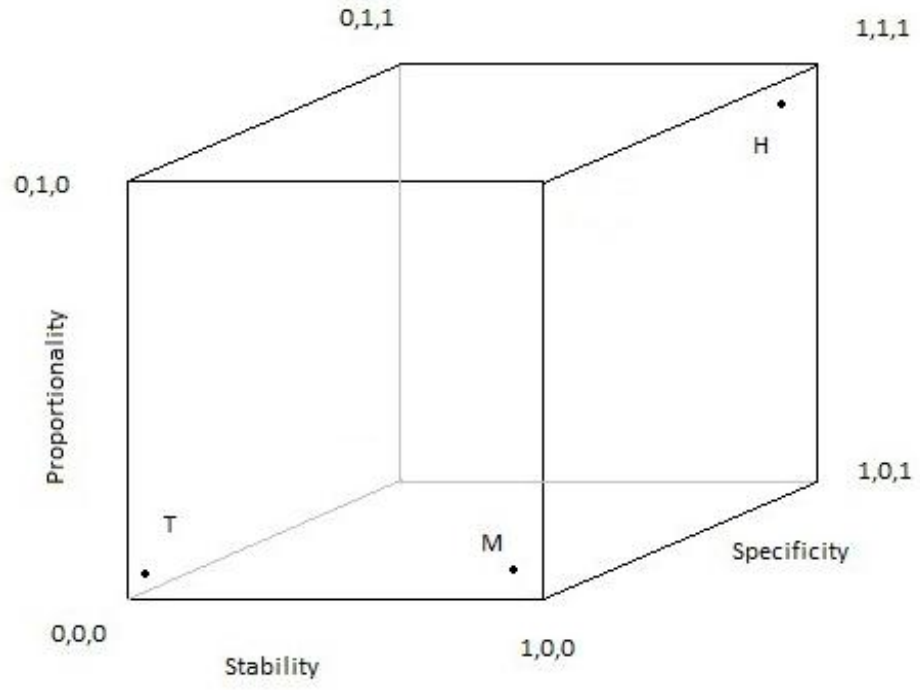



Figure 17: Causes of Huntington's Disease. T = temperature of the womb, M = cell membranes, H = huntingtin gene.

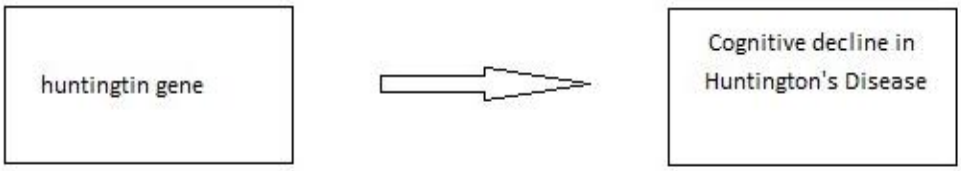



Figure 18: A Causal Explanatory Model of Huntington's Disease.

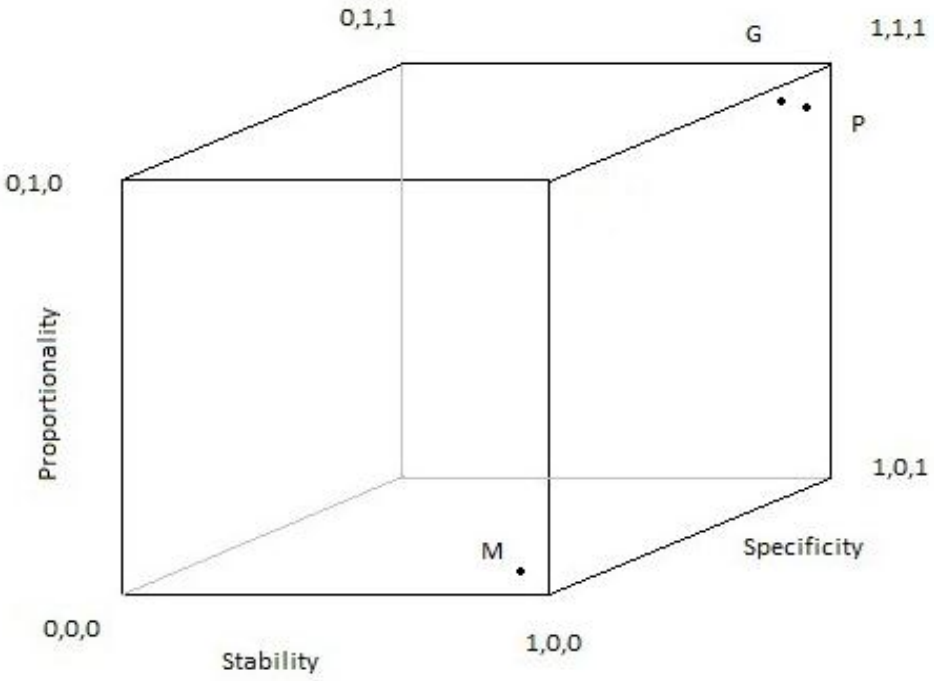



Figure 19: The Causes of Phenylketonuria. M = Membranes, P = Phenylalanine in the diet, G = Phenylketonuria gene.

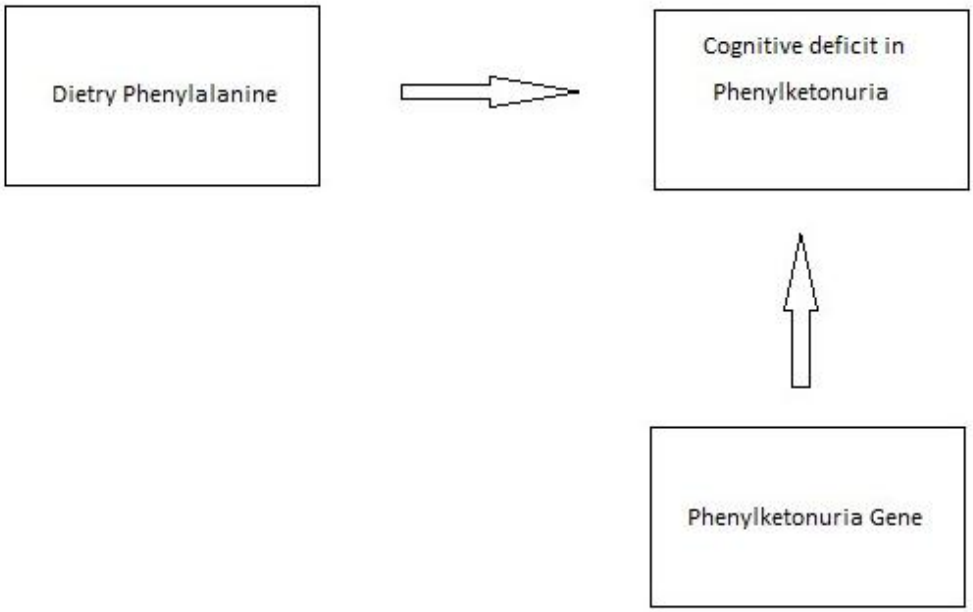



Figure 20 : A Causal Model of Phenylketonuria. Both environmental (cultural/dietary) causes and genetic causes are privileged causes of the disease and so must be included in the causal model.

# CHAPTER SEVEN: Theoretical Implications of the 4D Space

I have spent chapters four through six explaining and demonstrating a tool for analyzing the causes of some trait so that we can build an explanatory model focussing only on privileged causes. This shows that human technological innovations are sometimes privilged causes of human psychological traits. In this chapter I pick up on two theoretical concepts in biology, innateness and codes, and suggest how they may relate to spatial regions mapped out in the four-dimensional tool. Do some regions of the concept space correspond to these biological concepts?

## Innateness

In chapter two I examined the folk concept of innateness using the analysis that Mameli & Bateson have undertaken. This showed the folk concept to be wanting, and failing to connect to any particular useful and useable scientific concept. However, there may be a related concept of interest illustrated by the four-dimensional model. Once we determine the set of privileged causes for some particular trait, then we can ask the following questions. (1) Are the privileged causes stable? (2) Are they heritable (or better yet, are they inherited)? (3) Are they internal to the organism? To the degree that this is the case then the trait in question is of a particular kind, with a particular class of privileged causes. This seems to be the sort of trait we refer to when we make claims of innateness. However, if the privileged causes are not inherited, or not internal, or not stable[16], then we cannot make the claim that such a trait is anything like innate. Remember innateness is some approximation of invariantness and having an evolutionary selective history, though it is not clear how *exactly* we cash this term out when we use it in the folk sense. It does not seem to map to a single scientifically useful concept.

[16] A cause can be privileged and not stable in virtue of scoring highly on the other three dimensions.

For example, take the case of critical thinking skills. The privileged causes in the toy example we looked at were IQ, literacy and critical thinking courses. These causes may be stable in their relation to the effect (producing critical thinking skills), they may be heritable (literate parents tend to produce literate children), but they are not all internal to the organism. Critical thinking courses in particular are an external institution. But when we decompose IQ into its own causal analysis, we find that the privileged causes include environmental factors as demonstrated via the Flynn Effect. Literacy is heavily dependent on invented symbols and cultural practices as well. So critical thought as a trait is not of the class delineated by having stable, internal, and heritable causes constituting its set of privileged causes.

Huntington's disease on the other hand has a small set of privileged causes. This includes just the huntingtin gene in the analysis in chapter six. The huntingtin gene is stable, it is internal to the organism and it is heritable. I'm not arguing that Huntington's disease is 'innate'. I stand by the analysis of Mameli and Bateson outlined in chapter two. Innateness is not a useful concept, largely because it is unclear what we mean by the term. For example, in this case the disease is neither species-typical, nor a standard Darwinian adaptation. Both of these were argued to be elements of some folk conceptions of innateness. However, I am suggesting that there is a very definable concept, that where most of a trait's privileged causes are heritable, stable, and internal, that captures much of what we mean when we employ the term 'innate' in different contexts.

It may be objected that according to this analysis something turns out to be 'innate' or not 'innate' according to where we draw a line in the four-dimensional causal analysis tool, and that this is somewhat subjective. However, I suspect that there will be a lot of cases where the appropriate location for such a line is quite obvious. This will occur when there is an unpopulated gap in the causal analysis between the location of a cluster of causes or a cause scoring highly on many dimensions and a further cluster or scattering of causes not scoring highly on many dimensions. This is illustrated somewhat in figure 15, but more particularly in figures 17 and 19. Phenylketonuria is not a good candidate for 'innateness' according to this approach

because one out of two members of the obvious cluster of privileged causes (phenylalanine in the diet) is not internal to the organism, and is less obviously heritable.

The idea here is not that this is a fully worked out fix for the messy concept of innateness. But, that once we have analysed the causes of a trait, using the dimensional approach worked out in the last three chapters, we may then be in a position to more fully understand why a trait is generally considered innate or not innate in the folk sense. There may also be further kinds of traits defined by further clusters of privileged causal kinds, such as traits the privileged causes of which are *external*, heritable, and stable, for example. Such traits might also inform us about Human Nature.

## Developmental Codes and Representation

Sometimes talk about the development of traits is cashed out in talk of codes, or coding. Some causes of some traits are supposedly intructive, representative, or there is some rule-based translation mechanism at work converting causes into effects. We saw in chapter four that there is a project imparting informational and coding work to the genome in development, and also scepticism surrounding this project. But alleged developmental codes are not restricted to the genome. Other arguments and examples exist. Turner (2007) attempts to define an epigenetic code. He claims that modifications to the histone proteins around which DNA is wrapped constitute a code and have effects on which genes are active.

> 'An attempt is made to define how histone modifications operate as part of a *predictive* and heritable *epigenetic code* that *specifies* patterns of gene expression through differentiation and development' (p. 2, my emphasis).
>
> Furthermore, 'the increasing variety and interactive properties of histone modifications has led to the use of the terms "histone code" and "epigenetic code"' (p. 3).

In a developing system a developmental code is a code for some outcome of development. Codes are rules for converting information from one form to another. Biologists often talk about a distinction between permissive and instructive causes (see e.g. Gilbert 2003). A permissive cause is a necessary background condition for some effect to occur, an instructive cause bears some kind of imperative, plan, or program-like characteristics. We see now that some of this distinction is made salient in the four-dimensional space. Permissive elements will score highly on dependence and perhaps low as proportional causes in the four-dimensional space. Instructive causes on the other hand are generally what we mean by 'developmental codes'. Often it is argued that an imperative message or code in some developmental elements (such as DNA) is an important 'instructive' sub-category of cause.

In this section I will examine two ordinary codes, Caesar's Code and Morse Code, and deduce five criteria necessary for something being a code. We can then relate these criteria to the four-dimensional model and delineate a part of the concept space in which causes must be found in order to be considered as possible codes. Given that we are employing a parity of reasoning approach, we must approach all elements of the developmental matrix as if they could be developmental codes, and test each one. I will show that very few things, if any, can be seen as true developmental codes. And that even in cases where one can make an argument for a coding role for the causal factor, that explaining development in terms of coding adds nothing to the explanation offered by the four-dimensional model, and in some cases may confuse issues. Note that I am only focusing here on the processes of ontogeny, and that any talk of codes or explanation of inheritance across generations will not necessarily be captured by the present arguments.

The code employed by Julius Caesar to communicate with his generals allowed the conversion of coded messages into plain text by employing a rule where letters of the alphabet are frame-shifted three places. C becomes F, D becomes G and so on. Morse code is a more complex method for encoding the alphabet. Short combinations of dashes and dots in Morse code correspond to letters of the alphabet. But codes are also imperatives bearing the instructions, 'add letter α to the growing message'.

In Caesar's code changing a letter changes the corresponding letter in the message. In Morse code adding or deleting a dot or dash changes the letter in the message. In this sense both codes are specific causes of their messages. Small alterations of the code have corresponding small effects. Multiple states of the code C1, C2, C3... correspond to multiple states of the effect E1, E2, E3... This is our specificity criteria from chapter four.

These codes are also proportional. In Caesar's code one letter corresponds to one letter in the message. In Morse code one small cluster of dots and dashes corresponds to one letter in the message. Neither code nor translation is broader or narrower than the other. Codes are 'grained' such that identifiable units of the message are mapped to identifiable units of the form or representation produced. Letter to letter in Cesar's code, or small set of dots and dashes to letter in Morse code. Information is neither lost nor gained in the process of translating.

The question of causing and the question of coding are two separate issues. Some entity can cause an effect without coding for it. Legs cause us to walk, but we don't normally consider legs to code for walking. Genes are causes of psychological traits but this doesn't entail that they code for them. Codes on the other hand do need to be causes. Red lights code for traffic stopping and they also cause traffic to stop. Caesar's code causes intelligible messages to be extracted. Dolphins are obviously not a code for human eye colour because they play no role in causing human eye colour. The relationship between dolphins and human eye colour fails to satisfy (M), our minimal sense of a cause. Also, we should note that the vehicle (letters in Caesar's code, dots and dashes in Morse code) represents the outcome, but it is the rule (shift of three places) that is the code. Codes are systems of signs and meanings, and they necessitate adaptors for converting signs into meanings.

From this discussion it seems that we can begin to construct criteria for what it would take to be a developmental code. To be a developmental code it is not enough to be a 'gene' or a cultural technology, or some other class. Instead five criteria are required:

1. The entity in question must be a developmental cause (satisfying (M)) for there are no non-causal developmental codes. This criterion is necessary to omit cases like the dolphin example I have just outlined.

But satisfying (1) is far from sufficient to be a developmental code. Think of gravity. Gravity has causal effects in the development of probably every trait. But gravity alone does not seem to code for our traits.

2. The entity must be specific to the developmental outcome.

But causal specificity is insufficient for being a code. The ambient temperature has a causal, specific relationship to the molecular lattice of $H_2O$. Hot ambient temperature (C1) causes steam (E1), moderate temperature (C2) causes water (E2), and cold temperature (C3) causes ice (E3). But again, I contest that we don't ordinarily consider temperature to 'code' for molecular structure.

3. The entity must be proportional to the developmental outcome.

Tides cause high-water marks on beaches, larger tides cause higher high water marks, and there is tidy proportionality between the tide and the water-mark. High water marks are not caused by anything narrower or broader than the tides. But again, we don't consider tides as codes of high water marks. They are merely proportional, specific causes. Indeed, genes seem likely to fail a proportionality test for coding with respect to most traits other than perhaps amino acid sequences. And amino acid sequences don't seem to be developmental 'outcomes'.

The reason why we resist considering tides and similar examples codes is because they lack the quality of representation. Wheeler (2003) argues that coding talk is a species of representational explanation along the lines of instructions, blueprints, plans, specifications or a program. A code is a rule for converting information into another form or representation, not necessarily of the same type. So we need a fourth criterion:

4. The entity to be considered a code must represent the developmental outcome (i.e. be intentional).

DNA has a causal, specific, proportional relationship with some traits. We have already seen that. But in chapter four I argued that DNA does not carry semantic information. If semantic information is required for representation, and hence for coding, then DNA cannot code for developmental outcomes. There is a genetic code in one sense, this is the mapping rule between codons and amino acids. But this is not a code *for* some developmental outcome. It is criteria for this stronger sense of developmental code that I am trying to elucidate here.

What makes something a representation? Representations must represent *something*. Representations possess 'aboutness' or 'directedness' on something in the world. This is the quality of intentionality. Mental states are the paradigmatic cases of intentional things. Our mental states are 'about' things in the world. They have *content*. Furthermore something is only a representation if its function is to represent. Putnam (1981) gives the example of the ant causing a trail in the sand that just happens to look like a famous painting. But the trail in the sand does not represent the painting in and of itself. It only represents the painting when you or I or someone else comes along and uses the ant trail as a representation. This is one of the key reasons why DNA fails to represent the outcomes of development. It is simply not the function of DNA to do so. The function of DNA is to package itself for tranmission to the next generation and to be available as one of the ingredients in development. Objects in the world can acquire intentionality and therefore represent, thanks to semantic content *derived from* minds. So things that represent include mental states and artifacts, such as pictures, books, and language, which derive intentionality from minds.

So, representations of student cognitive developmental outcomes in the minds of teachers could satisfy these criteria. Teachers' mental states are causal, specific, proportional, representational things with respect to student learning. Teacher intention satisfies the criteria we have identified so far.

What about artifacts? First let me set this problem up with an illustrative example. When designing a house an architect represents his ideas with a plan or blueprint, the plan can be seen as a set of instructions for the construction of the house. The plan can then play a causal role in the construction of the house even if the architect is no longer present. The representations constituting the plan contain semantic information about the house because they are derived from the architect's mental representations, which uncontroversially represent the house. There is a special relationship between the architect's mental content and the proposed house. This is the intentional relationship. The plans are also characterized by this relationship. They possess intention derived from the architect's mental content. According to the criteria listed above the architect's plans code for the final structure of the house. This seems plausible. The architect's mental states (A), code for the plans (B), and the plans code for the house (C). The plans function to represent the house, they also cause the house, and they are specific and proportional to the house.

If agents are modifying cultural technologies with the intent of bringing about a particular cognitive device in their children, then there is a very strong sense in which they are representing the phenotypic outcome. The agents may also take steps to ensure that such an outcome occurs. Hence, redundancies and duplications creep into the strategy. We can attribute a representation here because the function of the agent's intent is to represent the cognitive outcome of the child. This occurs all the time in education. We have an analogous situation to the architect case. In the case of programming skills there is the teacher intention (A), to produce programming skills in the child, there is Storytelling Alice (B), which facilitates the outcome, there are student learning outcomes (C). But Storytelling Alice is less obviously a representation of the outcomes, than is the architect's plan. We need to ask, when something (A) is a code for some other thing (C) and there is an intermediary step (B), then does (B) also code for (C). We need to know if coding is transitive in the right way. There certainly exist counter-examples.

Suppose I wanted someone killed. I represent them being dead in my mind. This representation can be causal, it is proportional to their death, and it is specific to the mode of death, and so on. Even if we were to grant that my mental state codes for the death, then it is a stretch to claim that the bow and arrow, which I construct in order to facilitate the death, is a code. Something can function to bring about a coded outcome, without itself being a code. We may need to know in more detail whether Storytelling Alice represents cognitive outcomes in the right way.

What is clear is that anything that clearly represents the outcome of development is a good candidate for a developmental code. Genes, as I argued in chapter four, fail to represent anything, except perhaps amino acid sequences. But this does not mean in and of itself that there are no developmental codes. Cultural structures or mental representations could still be candidates.

However, there is a further problem. The criteria for coding as they stand are too broad. This is because nearly anything represented by a mind then becomes 'coded for'. If I have a representation of making a sandwich in my mind, and this causes me to make a specific sandwich, proportional to my representation of it. Then my mental state has coded for the sandwich. If we want to preserve the everyday notion that Caesar's code and Morse code are codes, and common mental states like representing a sandwich are not, then we need more in our criteria. The missing ingredient seems to be an interpretation rule.

5. Codes adhere to a rule of interpretation, which dictates how the represented entity is to be extracted from the representation.

This idea becomes clear if we think of a simple coding system. Take the system of red and green traffic lights. Red lights cause traffic to stop, and green lights cause it to flow. But a pile of rocks in the middle of the road also causes traffic to stop. The reason that red lights are a code rather than just a cause for traffic stopping is because there is a rule of interpretation, the code, which connects the sign, red light, with the outcome, stopped traffic. It is not that the photons of light falling on a driver's retina physically cause traffic to stop; the stopping is due to the meaning of

the sign. The driver of the car is an *adaptor*, converting the meaning of the sign, not just its physical effects, into an outcome.

The coding entity need not be stable, as many codes fail to code in the wrong context. For example a Chinese speaker who understands no Latin could not convert Caesar's code. The entity need not be enabling, the developmental outcome may still occur even if it is not coded. However, if some cause is specific and proportional, and furthermore it *is* stable or enabling, then it is clear that it palys an important role in development as a privileged cause, whether it codes or not. It is entirely contingent whether any actual causes of developmental outcomes satisfy these coding criteria.

Criteria (1)-(4) above are too broad an analysis of coding, and criteria (1)-(5) seem to leave us with few possibilities for developmental codes. Even mental states as causal entities don't obviously appear to follow a rule of interpretation. Genes fail criterion (4). The causal specificity between DNA codons and the primary structure of proteins can be called a code irrespective of whether it carries a message or instruction (Shea 2007). This does not imply that DNA has any semantic or intentional properties. But DNA in this sense certainly doesn't code for developmental outcomes. And a developmental code is a code for some outcome of development. Mental representations fail criterion (5). Caesar's Code, Morse Code, and other commonly employed communicative codes satisfy all five criteria. They are specific, proportional, representational, causes of their outcomes via a rule of interpretation. We can however assert that if anything is to qualify as a developmental code, it must appear in the shaded 'coding' region in the conceptual space model I have introduced as depicted in figure 21. Causes (notably genes) that fall outside this region simply are not codes for the effects of interest. Before even beginning to argue whether genes are representational or not, we immediately see that dysbindin cannot be a code for intelligence (figure 12) because it falls outside this region. The causal analysis helps us rule out many candidate causes as being codes for development.

This analysis of codes is intended to support the use of the four-dimensional model for causal explanation. If there are no developmental codes or if being a code is so

broad that lots of counterintuitive things qualify, then perhaps we ought to just drop talk of developmental codes altogether. The folk notion that genes code for traits is empty and problematic. Full explanation of development can be obtained through the analyses provided by the conceptual space model. Nothing is gained by talking of codes. And in fact such talk may lead to omissions of important causal factors in explanation. Causal analysis does all the work we need. There may be a sub-class of causes which we refer to as *causal representations*, but this is as far as we need to go. Even if there *were* codes we often wouldn't be as interested in these codes as we are in the actual privileged causes of the effect of interest. So the importance of codes and privileged causation comes apart.

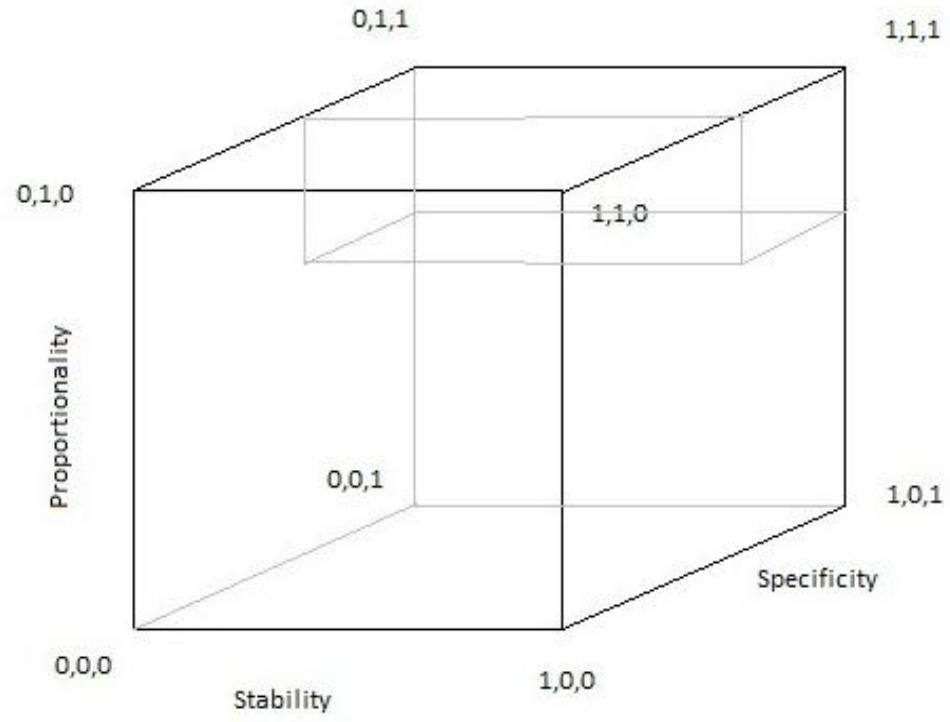



Figure 21: Codes. The smaller box represents the region of the causal concept space where a cause must be found if it is to have any chance of being considered as coding for some effect.

The criteria for being a developmental code are: (1) satisfy (M), (2) appear in the proportional and (3) specific region of the 4D space, (4) possess semantic information content about phenotypic outcome, and (5) satisfy an interpretation rule.

Testing the causes of critical thinking skills, we see that in the three-dimensional box (figure 15) student IQ and Critical Thinking Courses fall in the region of space necessary to be considered a code. However, student IQ does not possess the required representational content (nor is its function to represent the trait). Critical thinking courses on the other hand may be intentional, but there is not obviously a rule of interpretation linking the mental representation as 'code' and the outcome produced. Critical thinking courses are, however, heritable and, therefore, potentially of evolutionary significance. If there are selective pressures on critical thinking courses, then they will evolve over time leading to new critical thinking traits developing in subsequent generations. This sort of effect is the subject matter of the next and final chapter.

Formal education systems contain a wealth of cultural information and innovations to enhance teaching and learning. With developing awareness of these processes, lineages (humans) will be able to plan their teaching and engineer the education systems themselves. This requires a degree of knowledge of what is desirable, what has worked in the past, and what is likely to work in the future. Just like an architect's mental representation is part of the cause of a house, the mental representations and intentions of teachers, policy makers, and learners are part of the cause of certain traits. This has the potential to create a very powerful set of adaptive experiments.

Up to this point I have shown that some technolgies can shape our plastic brains during development and also integrate with brains to enhance human capabilities in the world. Furthermore, some human technologies can be identified as privileged causes of psychological traits. In chapter eight I move from looking at development, to looking at evolution. How does understanding all these developmental processes help our understanding of the evolution of the human mind?

# PART THREE:

## Evolution and the Sapient Paradox

# CHAPTER EIGHT: The Sapient Paradox and the Neanderthal Extinction

> 'Those living at Pushkari [20,000 years ago] are *Homo sapiens*, modern humans, anatomically and *mentally the same* as you and me' (Mithen 2003, p.10, my italics)

## Introduction

I have argued that cultural technologies are developmental causes of psychological traits. I have shown that the likely mechanism is an adaptation for neural plasticity and that psychological traits in some cases also functionally depend on technological aspects of culture. I then argued that cultural technologies can be privileged elements of the developmental matrix and should play prominent roles in the explanations of some psychological traits.

But the role of human technological innovation is not restricted to its developmental effects. This is because development plays an important role in evolution. Novel environments can cause novel psychological traits to develop, as we have seen. But then novel psychological traits cause the construction of new novel developmental environments (as we saw with the evolution of literacy). This ratcheting process can scaffold population-wide change in psychological traits over time. The arguments in chapter one suggest that the whole array of human psychology from affect, to intelligence, and from attention to memory, is manipulated by cultural variability. So these processes have important roles to play in the evolution of minds. The effect of computer games on emotion does not explain all of the changes in affect over recent decades, and the effect of minus signs and number lines doesn't explain all the changes in IQ. But the whole cultural suite taken together is a powerful driver of human cognition. This will have been the case ever since culture emerged in our evolutionary past.

In this chapter I apply this analysis to two problems in prehistory. The two prehistoric questions are, firstly, why did it take anatomically modern humans so long to become behaviourally modern? This is sometimes known as the 'sapient paradox'. And, secondly, given that Neanderthals were a lot like humans, why did they go extinct? In the process I am going to challenge Mithen's claim with which I opened this chapter. Our mental life has changed dramatically over recent prehistory.

I will argue that cultural-technological cognitive-developmental processes must have been occurring during these timeframes and that our plastic brains accommodated these influences by rewiring as well as integrating with cultural technologies in the world. Altered developmental environments provide scaffolding for novel concept acquisition, rewire the brain, and novel functional environments extend the functional apparatus of cognition. Minds have changed in many ways in both the take-off of sapiens' behaviour and in the Neanderthals' decay into extinction. I add at the outset that I am not attempting to give an actual sequence explanation of what actually happened in these two cases, rather I am outlining a framework which could be employed if someone were to seek the fine details.

## Approaching the Lineage Explanation

There are two types of questions that cognitive prehistorians can ask. One is what were people like at other times? What were people like at time A? at time B? These type (1) questions are descriptive questions. The second sort of question is a process question. *How* did people of type A become people of type B. Note that in answering type (2) questions we generate a lineage explanation and give an account of mechanism.

The evolution of the human mind is a long process that appears to advance in fits and starts (Donald 1991). This is not unexpected because this is the same pattern that we see in the biological evolution of organisms (Gould 2002). This is also the pattern we see in the macro scale changes in the earth's geology. In each case micro events feed up to macro changes and there is change over time. An important feature of such

change in these biological and geological cases is that change is ongoing. Organisms are evolving right here and now and the earth's geology is in constant flux. Every time there is an earthquake, or a landslide, or volcanic eruption the earth's geology changes. Charles Lyell appreciated this in 1830 when he published his: *Principles of Geology, Being an Attempt to Explain the Former Changes in the Earth's Surface by Reference to Causes Now in Operation.* In this he observed that we can explain how the Earth came to take its present form, by appealing to ongoing small geological events that we can study here and now. This is exactly the same sort of insight that led Darwin to his *Origin of Species*. It happens that there are indeed current ongoing small changes in human cognition from generation to generation. Effects such as the Flynn Effect on IQ (see chapter one for a full discussion) must surely hold some clues to the processes underlying cognitive evolution. Key drivers of the Flynn Effect seem to be a complex social media environment, and a particular sort of formalised education system. Changes to the wealth of public representations, and to the structure of education syllabi seem to drive current evolution of IQ. These changes are ongoing and occurring right now.

But we can look back further, remember the effects of the Cambridge tutorial system, then there are slide rules in the 19$^{th}$ Century (see Nickerson, 2005), mathematical notation systems in Mesopotamia, notched sticks at 30,000 years, number words, and the effects accumulate. We can expect such effects to be found since the origins of culture. The explanation for the rise of science, math and the sort of abstract thought that characterises high IQ scores turns out to be a series of small but important events that accumulate and alter pedagogic structures and developmental environments.

The Flynn Effect is a population-wide psychological phenomenon driven by ongoing change in developmental environments. There have been other important population-wide psychological changes in the past. Lyell's phrase, '*Causes Now in Operation,'* suggests explanations for these phenomena. It is the change from one context to another that will provide our lineage explanation. I have suggested the elements of an explanation of the rise of scientific thought, and IQ effects. Similar

frameworks ought to apply to other significant cognitive emergences such as: language, literacy, aesthetic and representational art, self-reflexive awareness, and so on. Much of the explanation of these changes will be a changing pedagogic and media context for development.

I am not trying to *give* a lineage explanation. I am trying to explain how one might be constructed. An understanding of these contemporary processes *lessens the puzzle* of the sapient paradox. First, let's look more closely at the two problems I have introduced.

## Problem 1: The Sapient Paradox

There are two interesting issues troubling those who seek to explain the evolution of 'modern' humans. The first is the long time-lag (perhaps 150,000 years) between the appearance of anatomically modern humans and that of behaviourally modern humans (Renfrew's 'sapient paradox', 2008, chapter 5). The second is the unexplained demise of the Neanderthals, who in many respects were very similar to the successful *Homo sapiens*.

Genetic differences between contemporary human populations seem to be small and given that there was a radiation out of Africa approximately 60,000 years ago, which seeded all current global populations, then this suggests that the human genome has been relatively stable since then (Renfrew 2008). However, a study by Williamson et al. (2007) used statistical methods and found that up to ten percent of the human genome may have been affected by natural selection subsequent to the out of Africa radiation. Genes involved in skin pigmentation, lactose digestion, and the immune system, among others, have shown evolution. The gene for a single protein necessary for lactose digestion appears to have arisen in European farming populations over a period of just a few thousand years (Richerson & Boyd 2005).

In addition to this, Laland (2008) reviews evidence that recent selective pressures have shaped some genes involved in brain development, such as genes for serotonin

transporters and proteins associated with synapses. However, it seems unlikely that gene networks coding for specific cognitive devices could have emerged (and emerged in such numbers) over this evolutionary moment. What is more likely is that this is ongoing selection for efficiency and optimization of the system, quantitative, rather than qualitative improvements. The gene for a lactase enzyme could arise rapidly and effectively because it takes only one enzyme, and hence one gene, to facilitate the digestion of lactose. Modifications to one enzyme are unlikely to have effects on specific cognitive devices. Rather global gains or losses of degrees are probable as I have suggested in chapter two.

So the conclusion seems to be that even if there has been some genetic change over this timeframe, then it does not account for the cultural and behavioural diversity that we see. The genetic differences between human populations seem to be small in relation to the observed behavioural differences. Ongoing genetic changes in the sapiens lineage during the last 200,000 years are of relatively minor importance for behavioural change when compared to the gradual, gene-culture co-evolutionary process (see Laland 2008, Laland et al. 2010, Williamson et al. 2007, & Sterelny 2011 for examples). Culture has driven most of recent human evolution.

Why, if the genetic frame was relatively fixed by around 100,000-200,000 years ago did it take another 40,000 or more years for anything like human behavioural modernity to emerge no more than 60,000 years ago. Furthermore, why did it take *another* 50,000 years for really fancy behaviours like the invention of writing systems, long distance trade, and states to emerge in the Holocene (i.e. the last 10,000 years)? It is possible, I concede, that the sort of ongoing genetic changes I have just mentioned account for some of these delays, but given the arguments in this thesis to date, I suggest that there are several other legitimate answers to this question. It could be that:

1. There were important social transitions that were required for the emergence of really modern human behaviour.

2. There were ecological factors not yet in place in order for modern human behaviour to emerge.
3. Truly modern human cognition requires the accumulation of certain cultural technologies in order for it to emerge developmentally.

I am not ruling out some role for genetic evolution subsequent to *Homo sapiens* attaining anatomical modernity. Such 'invisible' (to the paleontological record) evolution may be relevant. But the role for genes becomes less and less likely as we approach changes in cognition and behaviour in the Holocene, such as symbolic proliferation and the origins of writing and mathematics. Furthermore, looking at contingent brute events does some part of the explanatory work. For example, social stratification and strong central government emerged as a result of a switch from dry to wet rice production in Madagascar (Linton, 1933). The small amount of very wet land at the bottom of a valley became especially valuable and property rights and a stratified family structure emerged (Trigger 1968). This in part explains the rise of cities and as we have seen in chapter one IQ in city contexts tends to be higher than that in rural contexts, there are rural/urban psychological differences. But this is far from the whole story. The explanation of human behavioural take-off is likely to be rich and complex.

One deflationary way to account for this problem is merely to suggest that the materials and methods needed time to accumulate. A lot of behaviour depends on accumulated technological mastery. For example something like reliably preparing cooked meat depends on a sequence of steps in the mastery of fire, each of which is difficult or impossible to invent before the ones preceding it. Building aqueducts requires that a host of more primitive technologies such as mathematics and brickwork be mastered first, and in sequence. The smelting of copper (first performed about 8000 BC in Turkey) required fired-clay technology first, in order that a ceramic crucible could be buried beneath a hearth and oxygen fed to the fire through blow tubes. Of course the dynamics of recombining precursor technologies to produce new innovations are interesting in their own right and we see the exponential take-off of a

suite of innovation once enough technological know-how is in place. According to this account, we first got smart, then bit by bit invented fancy technology.

On the other hand, much of what is fancy about contemporary humans is not obviously based in accumulating material culture. For example cognitive skills like syllogistic reasoning, or mathematical capability, don't seem to depend on such a long lag of accumulated material culture. Why do these not appear, and indeed in written forms, much earlier than they do, given that sticks for making marks on rocks were certainly available to the first anatomical moderns? Indeed, the first cognitive artefact that we know of created for mathematical purposes were 29 notches on a baboon fibula just 37,000 years ago (Mercier 2006). Similarly, Marshack (1972) estimates that French bone artifacts acted as lunar calendars as early as 30,000 years ago. So tallying, the correspondence of marks with units of quantity, seems to be approximately this old.

The key to explaining the sapient paradox lies in accounting for the cognitive skills that were required to drive behavioural modernity, and in realizing that these cognitive skills did not appear at the same time as anatomically modern humans. The first anatomically modern humans would have had an anatomically modern brain but the brain is not cognitively sophisticated on its own. Cognitive devices need to be constructed via developmental processes and integrated with supporting media. If developmental environments weren't building the right kind of cognition, then fancy modern behaviour would not exist, no matter how modern the genetic and anatomical resources of the species.

Renfrew (2008) describes a transition in human cognitive evolution, which comes around 40-60,000 years ago. Around this time there begins an increasing mismatch between the pace of genetic change (which continues at a slow steady rate) and the pace of behavioural change (which accelerates dramatically). We start seeing lots of behavioural change with very little genetic change (see figure 22). If the behaviour is a reflection of cognitive change, then there has to have been a 'tectonic' (or

constructive) phase of human cognitive evolution[17]. Cognitive nativists (such as proponents of Evolutionary Psychology, see chapter two) seem to be discussing human cognition that had evolved before this tectonic phase, whereas context-dependent constructivist developmental explanations, become much more relevant after this time as the technological context begins to rapidly accumulate.

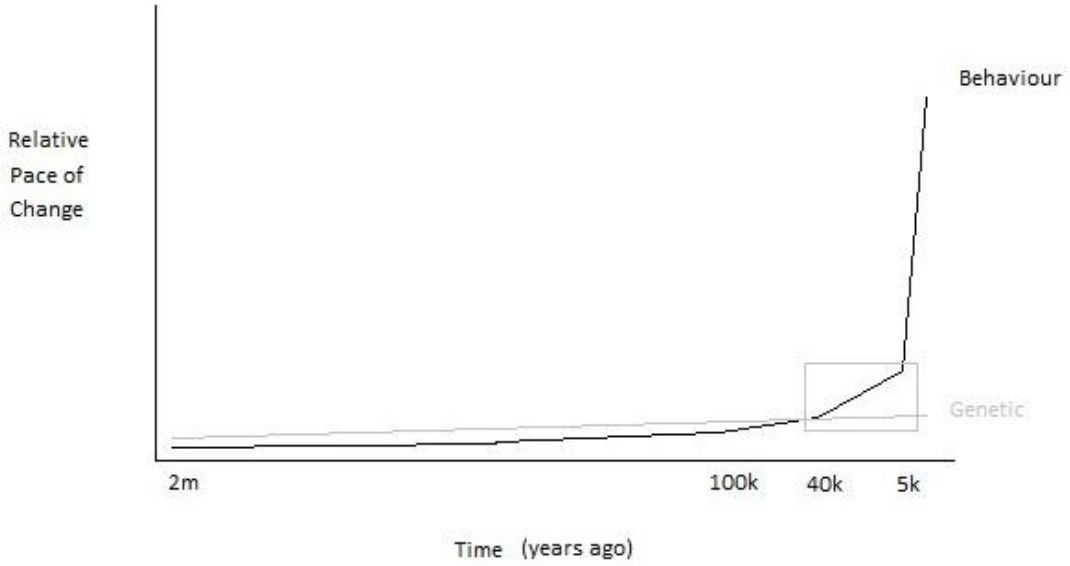



Figure 22: Pace of Genetic and Behavioural Change in Homo sapiens. Note that there is a period where the acceleration of behavioural change is dramatic. This is to be contrasted with the relatively steady rate of genetic change. Hence, as we approach the present, behavioural change is less and less likely to be explained by genetic changes. I argue instead that cultural technological changes are largely the cause.

Moving forward even further in time, we must emphasize the difference between humans today and those (pre-agricultural and pre-literate) of even 15,000 yrs ago. It seems that distinctively human cognition may have emerged very recently in the piece. This is the thesis of *human recency*. I argue that the deep structure of our cognition (e.g. language capability, a plastic neural cortex, a set of attentional biases, all the paraphernalia of general purpose learning) was in place a very long time ago, yet we had to wait until certain contextual conditions obtained before *humans like us* actually emerged through development. The upshot of this is that development has

[17] It is interesting to note the geological connotations of 'tectonic' in Renfrew's account, given what I have just said regarding Lyell's 19th century book.

sustained evolution for human cognition over the last handfuls of millennia. The flip side of this is that many novel cognitive devices could in theory disappear 'overnight' should contexts change again.

Overall, this means that if we are trying to find out about human minds, then nativist approaches may have blind spots as we begin to approach questions pertaining to more recent cognition. This is a *temporal constraint* on method. The methods of Evolutionary Psychology will not work as the timeframe of interest approaches the Holocene boundary and into more recent times. It is clear that the emergence of cognitive representations for reading and driving cannot have an evolutionary genetic explanation. I should note here, however, that the creation of these activities may alter selection on genes for neural development and change experience-dependent gene activation patterns. It is conceivable that efficiencies in reaction speed or alcohol tolerance may accrue after many generations of car driving. On the other hand more ancient cognitive drives like mate selection and landscape preferences may well have a genetic foundation. I can be agnostic on such claims for now. The more recent a phenomenon, the more likely it is to have a constructivist explanation. Given this, it seems that strict nativism is ill-equipped to answer questions about the cognitive evolution of ritual, art, material symbols, agriculture, value, the sacred, long-distance trade and human rights. The more recent our domain of interest the more that constructivism is going to be an important explanatory tool. I am not endorsing a pulse or burst model of human behavioural take-off here, rather I am illustrating that the relevant explanations are going to be ones of degrees.

A second constraint on the methodology of cognitive prehistory is the *embeddedness constraint*. Organisms exist in niches and niches (seen as relationships between organisms and their environments) shape both selective regimes and also development. Radical changes in niches, as clearly occurred in human populations over the last 40,000 years with the advent of art, numerical representation, literacy and distinctively modern culture, are likely to cause radical changes in development of the individual minds embedded in these niches. Radical changes in development can lead to radical changes in future niches thanks to the constructive efforts of

individuals. *We must understand development in order to understand evolutionarily recent processes.* The less genetic change that is occurring, relative to the weight of niche change, then the more that temporal change in developmental environments will explain process questions. Furthermore, we could reasonably argue that *most* of human cognitive evolution, or at least a significant chunk of it has occurred in this recent timeframe. Again, explaining why and how this happened explains the sapient paradox. I will give the framework for such an account shortly.

## Problem 2: The Neanderthal Extinction

The next problem of prehistory is this, Neanderthals were probably a lot like us and yet they went extinct. There are many reasons to believe that Neanderthals possessed a cognitive capacity roughly equal to that of early humans. This includes a large cerebral cortex, a larynx indicative of language capability, and tool-making behaviour necessitating essentially modern cognition. The Neanderthal brain at birth was similar to that of modern humans, and their brains grew to over 1600cc (larger than ours), yet Neanderthals remained dumb nephews in comparison to the behavioural take-off of *Homo sapiens*, having apparently never mastered symbolic technologies despite contact with symbolic and artistically capable *Homo sapiens* (Finlayson 2004). Furthermore, Neanderthals' last common ancestor with *Homo sapiens* was approximately 500-700,000 years ago, so Neanderthal's cognition should be considerably more human-like than that of chimps (with whom we share a last common ancestor an order of magnitude older again). And there is reason to believe that encultured (but never wild-type) chimps are able to pick up symbol use (see chapter one) and may have a 'language ready brain' (Taglialatela et al. 2008, 2009).

It is entirely possible that all the pre-requisite capacities for modern language (or indeed any cognitive trait) be evolved and available to an organism and yet there is no guarantee that an individual or a population of such organisms exhibit the trait. The right integration and developmental construction of these devices may simply not occur for contextual reasons. This seems to be demonstrated by the Moghul Emperor Akbar's deprivation experiments. This is largely because of the cognitive re-

tooling processes that I have outlined in chapter two (see also Lock & Peters 1996). So the essential capacities, the neural suite, of Neanderthals could be similar to that of humans, while their cognitive abilities are very different because their final configuration has been constructed differently in development.

If we are assuming biological comparability, then the explanation for the Neanderthal extinction must be contextual. It is not that they lacked cognitive potential, or were too stupid to adapt to a changing environment. A popular account is that ecological context changed dramatically forcing the Neanderthals to extinction (Finlayson 2009). But there are also many reasons for supposing that Neanderthals, at least in their last millennia, lacked the appropriate developmental context for developing human-like minds. What Neanderthals were probably missing was the appropriate scaffolding in their environment to turn their human-like developmental resources into the same sort of minds that *Homo sapiens* possessed. Just like wild-type chimps lack the scaffolding to develop symbolic communication like their encultured contemporaries. The argument of this thesis is that technological culture was an important part of this cognitive developmental environment, though learning forms and inheritance mechanisms are crucial, too, as we shall see.

My overall claim is that variation in the technological context of humans and Neanderthals explains, to some approximation, the different evolutionary trajectories that these two species took. The argument will be that both species constructed a technological niche, which had downstream effects on the development of subsequent generations. I will also argue that cultural technology sometimes evolves by Darwinian natural selection, and that adaptive cultural learnings become assimilated into formal educational practices over time. These three processes, when they work well together explain protracted behavioural take-off without appeal to genetic change, and when they fail these processes (or their lack) explain disastrous population collapse.

## Niche Construction, Hoyle and Evolving Technological Systems

Three theoretical approaches underscore the importance of human technological innovations for human evolution. The first is downstream niche construction, which I introduced in chapter three, and which I will elaborate on here. The second is analysis of the evolvability conditions. In this case the features that technological culture must exhibit in order to evolve by Darwinian means. I will take up Sterelny's account of evolvability using his 'Hoyle' criteria. This will allow us to test cultural technological innovation for evolvabililty. Finally we have theories that describe the adaptive response of the genome over generations as organisms learn adaptive behaviours ontogenetically. I will discuss a cultural analogue of this process, cultural Baldwinization, and explain how it can occur and cause important cultural technological innovation to become part of the package of cultural inheritance reliably passed from one generation to the next. I take each of these concepts in turn and then move on to resolve the two puzzles of prehistory that I have just introduced.

### Downstream Niche Construction

As well as supporting cognitive function by extending minds (see chapter three), niche construction activities can have downstream consequences and hence there can be an ecological inheritance. For example, humans typically live with their parents and benefit from their ecological engineering (Sterelny 2003). This may include shelter, educational resources or a productive agricultural environment. This ecological engineering is often cumulative. We saw in chapter three that organisms (beavers, meerkats, humans) actively construct their environments and that this has consequences for development and function. However, niche construction also has implications for evolution. A lot of niche construction is transmission of information *acquired* by the parent (Stotz 2010) (perhaps from the previous generation). But a lot of niche construction is information *created* by the parent. Each generation can add to and improve the ecological inheritance. This often includes adding to the accumulating knowledge-base of the social group. The Cambridge University tutorial example I gave in chapter three is a nice instance of this sort of process. Much of

human niche construction is this sort of 'epistemic engineering', which alters the informational character of the environment. This can transform problem spaces for learning (by transforming the developmental environment) or for functioning (by providing external supports for cognition) in ways that aid thinking and reasoning (Clark 2008, Sterelny 2003).

As I've already mentioned with the emergence of the lactase gene in farming populations, and potentially with the car-driving example, niche construction activities can alter selective environments so that different selective pressures on genes arise. Even if the brain develops in constructive fashion and cognitive devices are not genetically pre-specified, then adjustments to genes can still result in growth, development, and functional, efficiencies. Developmental niche construction will always entail the possibility of selective niche construction too[18].

However niches require resources and effort to sustain them, particularly if technology needs to be continually reproduced or maintained. Stotz writes that such 'exogenetic legacies demand continuous behavioural effort to maintain their value', (Stotz 2010, p. 491). In light of this we see that developmental niches and the affordances that they provide are *not* inherited as surely as genes. This is especially true with technological culture that requires human input for its maintenance. If the conditions that sustain the technology or niche begin to erode, then the developmental effects of the niche will erode too. My children will only inherit my house as shelter so long as I work to maintain its upkeep. If I am otherwise occupied, say in herding cattle on dry barren land, and struggling to provide food, then this potential ecological inheritance is lost. We will see shortly that this feature of ecological inheritance has particular relevance for explaining the Neanderthal extinction.

We can distinguish between education and mere development. Some niche construction is *intended* by members of the previous generation to bring about

[18] This entailment is strong. There are not two or more versions of niche construction (selective and developmental) as some authors (e.g. Stotz 2010) have suggested. Such distinctions give the appearance of controversy where none exists (or none ought to exist).

certain bodies of knowledge. This is called education. Many examples of intentional, formal, education abound, one need only look to the software programs I have already outlined, which help to bring about critical thinking, programming skills, or literacy. Other niche construction *just happens* to have certain developmental effects on offspring (or indeed the generation which created it) these are unintended developmental effects. An example of this might be the way that complex media contributes to rising IQ scores, or changes to self-representation are brought about by the use of virtual worlds.

So niche construction activities can sometimes have important adaptive developmental effects when they are sustained. But how are they sustained? I will give two examples now, one where developmental environments are sustained because cultural technology evolves adaptations. The second is where constructed niches are sustained because information that was once learned by trial and error by individuals becomes assimilated into the formal epistemic inheritance of the next generation through education.

### Hoyle and Evolvability

If there was some innovation, X, that altered the functional or developmental context of human brains, then we could see psychological trait change in a population. If X causes psychology that lowers fitness, then we would expect X to vanish from the population all else being equal. For example, cults tend to arise spontaneously, indoctrinate their members with bizarre beliefs and desires and then vanish rapidly because the psychology they create tends to be maladaptive. We could even see a sequence of different innovations, X, Y, Z... over time as new spontaneous innovations arise and vanish. However, if X enhances fitness, then we would expect X to proliferate, all else being equal. For example a writing system allows users to make lists and hence remember much more than they otherwise could. Such innovations tend to persist. Furthermore, if X is *evolvable*, and the effect of X is not at a fitness optimum, then we would expect X to evolve over time toward a fitness optimum, if there is one.

Niche construction activities aside, environmental conditions are not shaped by selection, so we don't ordinarily talk about an environment for phenotype X with respect to natural selection[19]. Environments are not seen to be evolvable in the same way that genes are. But why are genes evolvable? Sterelny (2001) offers some suggestions. If Hoyle and Co. (evolution engineers) were employed to construct an evolvable system what features would it need? Firstly, an inheritance mechanism is of prime importance. What is inherited may be called the replicators, and the system needs to have an anti-outlaw mechanism to prevent maverick replicators going it alone. Next, there must be stability from generation to generation. Inheritance must be of high fidelity, with redundancies built in to the system and a stable context for the inheritance mechanism to operate in. Finally, the system must generate variety. But, there has to be a smooth map from the nature of the replicator to the organization it causes. The variety must be modular in nature, and be especially open to duplication. This rough characterization of the Hoyle criteria will do for now. I will assume that the Hoyle criteria are what is required for evolvability. Genes, it is argued, possess almost a 'Full Hoyle' (note that representation or information are not part of the Hoyle criteria), and therefore genes can sustain an evolvable system. It is on the basis of these criteria that genes have an important relationship to evolutionary processes.

In the case of genes there is argued to be an appropriate Hoyle relation between genetic replicators and phenotypic organization. Is there a Hoyle relation between cultural technologies and psychological traits? It has become apparent that changes in human technologies can change the developmental context for cognition. We can see different cognition in different contexts. But can cultural technological change *evolve* new cognition?

[19] Kitcher accepts that there are 'environments for' as well as 'genes for' traits. But when challenged to say that environments 'code' for traits he dismisses talk of genetic coding as a rhetorical flourish that plays no part in explanation (Godfrey-Smith 1999). Hopefully my arguments in chapter four against the notion of genetic information make sense of this claim.

It is perfectly clear that much culture is not Hoyley. Lots of culture is not reliably inherited, it is the innovation of individual agents and it passes with them (of course this does not mean it can't influence cognitive development, just that it won't cause cumulative evolution of psychological traits). Other culture *is* inherited, but is so transformed as to be almost unrecognizable, such fashions, trends, and general contingencies pervade cultures. Other culture *is* inherited, and reproduced reliably. Writing systems are a good example of this.

I am arguing that some cultural technologies, in particular some educational and symbolic systems flow in Hoyley fashion through cultural lineages. Some educational technologies (in particular) scaffold the socialization, development, and learning of students, and so play an important role in generating those who then make further educational technologies and pass them on to others. Educational technologies often enhance the biological fitness of agents. Think of master-apprentice relationships in stone tool knapping. Students often learn from more than one source and so we see duplication of the educational source. This means that one teacher can experiment with innovation without risking too much. This is particularly apparent in the Cambridge tutorial system's evolution. The tutors were free to innovate and yet the students would continue to learn physics from the university itself. The inclusion of different symbol systems, different disciplines, different teaching regimes, artifacts, other technologies, and so on, makes education a highly modular and mutable system. There is great stability of some education institutions over time. There is material overlap that can scaffold reproduction of the cultural technology. There are redundancies in the systems, and the steady shift in human IQ over the twentieth century suggests a smooth map between education technologies (or the learning environment) and cognitive effects. Indeed, educational technologies as a replicator set seem to be a prime candidate for Hoyleyness if anything is. Cultural inheritance is not effected through a homogenous channel, and some channels are more Hoyley than others.

If we think about the effect that cultural technologies have on the developing mind, such as the effect that playing a violent game like 'Grand Theft Auto', or other role-

playing war scenarios, has on aggression, then we shed light on an important relationship between cultural technologies and cognition. I am suggesting that if the Hoyle criteria (or something similar) are the right way to characterize what it means to be an evolvable system, then cultural technologies may be important in the evolution, and adaptation, not just the development of the mind. Furthermore, if culture is an inheritance system that allows us to speak of culture 'for' some cognitive phenotype, then culture is not just materially important in extending the sapient mind, and is not just developmentally important as a changing milieu, but is evolutionarily important in explaining complex cognitive adaptation. I now want to give one concrete example of this process in action. I turn to the evolution of writing systems.

### Case Study: The Evolution of Writing Systems

Many authors argue that human culture, at least in many instances, evolves by Darwinian natural selection (Cavelli-Sforza & Feldman 1981; Boyd & Richerson 1985; Richerson & Boyd 2005; Dennett 1995; Mesoudi et al. 2006; Shennan 2009). Those who study writing systems have argued whether or not writing is a case in point. Some fiercely reject the idea (e.g. Barton & Hamilton 1996, Houston 2004). Others have been more sympathetic (e.g. Trigger 2004; Changizi and Shimojo 2005; Skelton 2008). At times debate has been ill conceived because of confusion over what the notion of a writing system 'evolving' actually entails. There are at least three distinct theses intended when it is suggested that writing systems evolved. First, it may merely be meant that culture changes over time. This sort of evolution is surely trivially true (this still leaves culture as an important developmental cause as I have argued to this point). Second, it might be meant that some goal-oriented or progressive process is at work. Third, and as it is conceived in this chapter, cultural evolution is at least in some instances a Darwinian process. That is, culture evolves when there is descent of cultural traits, with modification, modulated by a process of selection.

Writing systems seem to have a good chance of sastisfying the Hoyle conditions. Inheritance can be demonstrated using phylogenetic systematics (Skelton 2008, see figure 23). Linear B was a writing system used to keep economic records on clay tablets on the Greek mainland and on Crete between 1450 and 1200 BCE. Skelton notes that palaeographic techniques, in which differences in sign form are used to judge how closely two writing traditions are related, have been employed in discussions of the evolution of linear B. However Skelton then adapts phylogenetic systematics to evaluate this paleographic evidence. Taking the nuances of scribal hands as data, and applying the algorithms of phylogenetic systematics he infers the relations between scribal hands. Overall Skelton finds that when the data are analysed with criterion for finding the optimally parsimonious phylogenetic tree, the tree produced is largely consistent with the historical context of Linear B. This includes lending support to the existing theory that two scribal hands found at Pylos pre-date other materials from that site.

The scribal hands in Skelton's analysis show variety. The fidelity of inheritance is high, writing systems are reproduced with great accuracy and are stable across time. In some cases more than ten per cent of cunieform texts are lexical texts, or instructions for future scribes to learn and use the system accurately (Cooper 2004). Great attention is paid to reproducing the systems correctly, as is evidenced by the attention to detail teachers pay to students' handwriting in junior classes. Ensuring that a script survives necessitates teachers and apprentices (Houston 2004). This requires a high bandwidth of social inheritance, but writing also massively enhances bandwidth, so we see a ratchet effect. Writing down writing systems enhances stability. Past markings can act as templates for future markings. Writing systems are modular, elements can change or be assembled piecemeal. Written numerical notations are independent of written logograms, and these in turn are independent of the graphical representation of speech, or diagrams, or punctuation marks. Finally, there is a smooth map between many different forms of writing systems and the abilities that users of the systems can possess (as I have argued when discussing the functional extension of information processing loops in chapter three). So there

seems to be a case for at least one cultural technology (writing) satisfying at least one set of criteria for evolvability (Hoyleyness).

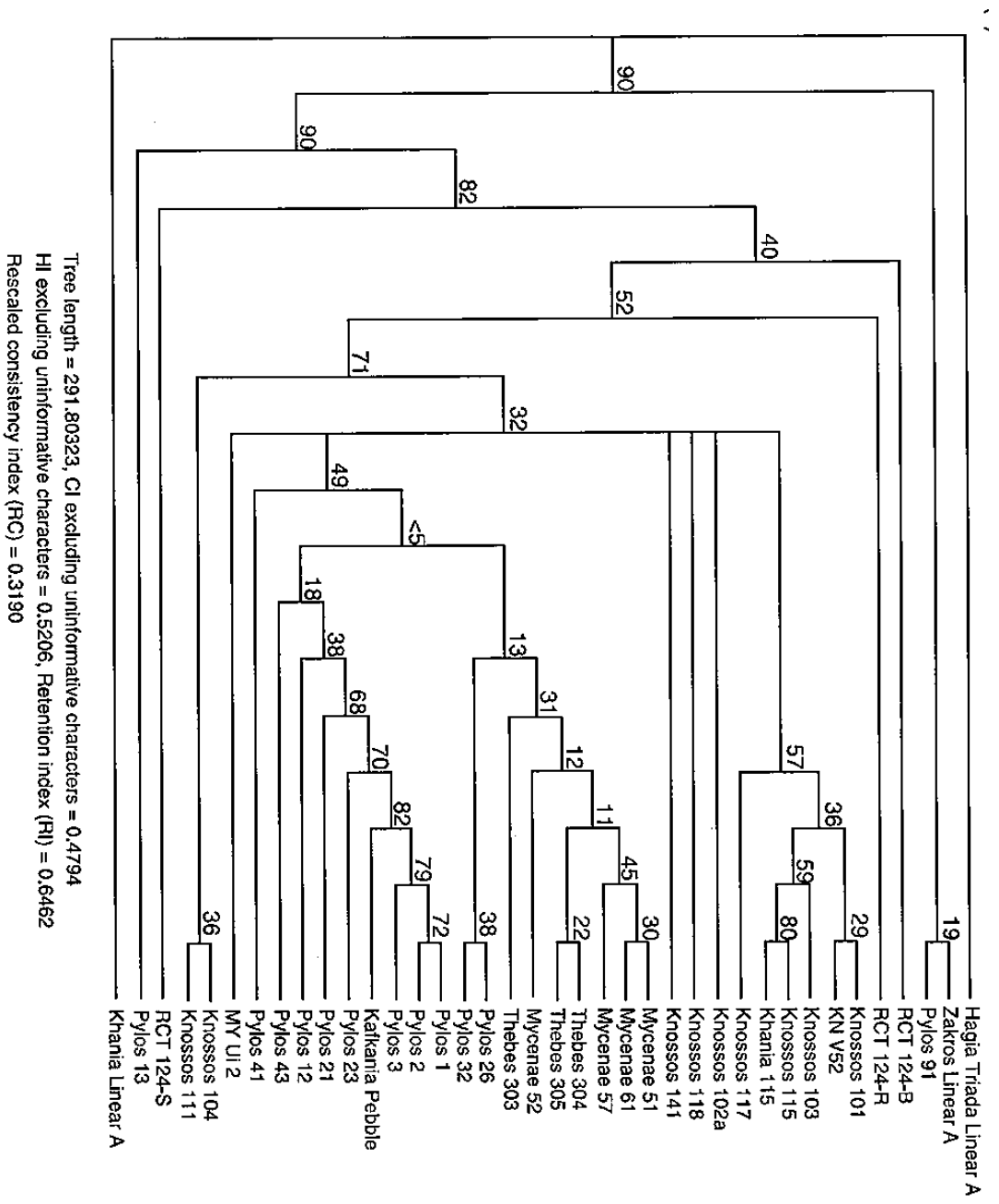



Figure 23: Phylogeny of Linear B (Skelton 2008)

### Cultural 'Genetic' Assimilation

Humans construct developmental niches. These can in turn be selective on genes. Human technology, which forms a part of the developmental niche, can evolve in

Hoyley fashion. But, humans also learn to figure things out themselves, and learn adaptive behaviours by trial and error. If populations happen upon good tricks via this method then two things are likely to happen. There could be genetic selection for features of brains that enhance the tendency to acquire the good trick. And the good trick is likely to be assimilated over time into the cultural inheritance of that population, so that individuals won't need to waste energy learning it for themselves.

There may be some technology, which is learnable on one's own (say, the innovation of notching a stick to track the full moon), which enhances psychological trait fitness, such as memory. In an innovating population we would expect some members of the population to learn this technology. If the technological innovation really does enhance psychological trait fitness, then we would expect, over time, the technology to become a part of cultural inheritance.

Isaac Newton experimented with cultural technology and composed a set of technical signs, symbols and rules for mathematically describing the world. Subsequent to Newton, this set of signs and rules became a standard component of the school curriculum and has been taught to a great many students. It has in a sense become part of the cultural DNA of our developmental environments. This instance is a cultural analogue of the general evolutionary process known as genetic assimilation, of which the Baldwin effect is a well known example.

Genetic assimilation occurs when a population is initially occupying one environment, although there is an unexpressed genetic capacity for plasticity. When the environment changes the pre-existing capacity for plasticity allows the population to adapt through novel phenotypic expression without any genetic change. The new phenotype may then become genetically fixed or assimilated (Pigliucci et al. 2006). I will give two specific instances of this general effect. First the Baldwin effect, and then Avital and Jablonka's (2000) 'assimilate and stretch' model.

The Baldwin effect (see West-Eberhard 2003, Sterelny 2004) occurs when a population encounters a novel adaptive problem and initially overcomes it thanks to

adaptive phenotypic plasticity, often in the form of learning to solve the adaptive problem during ontogeny. Efficient development of the successful trait is then selected for and over time development of the trait with less experience or environmental input emerges because of genetic evolution favouring ease of acquisition of the trait. At the limit the trait develops in the absence of experience. This process could play a role in shifting privileged causation from external to internal factors hence increasing the degree to which a cognitive trait has internal, stable, heritable, privileged causes (see chapter seven).

Avital and Jablonka (2000), and Jablonka and Lamb (2005a), characterize a process they label 'assimilate and stretch'. In these cases both the trait and the environment are transformed, perhaps continuously. The environment may change first, and as above, the trait changes through pre-existing plasticity mechanisms. But then a related trait, for example a correlated behaviour, can be conditionally expressed with much higher probability. In some cases the genetic assimilation of a behaviour, which would have been laborious to learn by individual trial and error, leaves resources available for learning still further adaptive phenotypes. These in turn can also be assimilated genetically.

If cultures are evolvable, they are also open to Baldwinian and 'assimilate and stretch' processes. If a trait is successful and it can be more efficiently taught than learned, then there will be selection for more efficient teaching of the trait. At the limit the trait is learned from teachers (perhaps in complex formal fashion complete with structured exercises) without unstructured experimentation by the agent alone. For example, in the 1970s and 1980s computer skills were largely learned by innovative individual trial and error. Some mix of inputs, largely drawing on informal learning with only a minor contribution from formal learning accounted for the development of computer skills. Over time computer skills have proved to be adaptive in our current environment and over time the mix of inputs, which generate computer skills has shifted to include more and more formal teaching to the point where almost all primary school children get instruction in computer use. This could be seen as *cultural* as opposed to genetic Baldwinism.

Remember the IQ rises discussed in chapter one. Over the 20th century people have generally become less cognitively linked to the concrete and more flexible and able to manipulate abstract ideas. Importantly technologies like teaching, textbooks, syllabi, and software applications are all cultural objects that are inherited and open to natural selection. Over time such technologies that cause useful cognitive skills to develop ought to be selected for and refined. It doesn't matter whether or not increased fluid cognitive ability enhances human biological fitness, if it is something we value as a society then the mechanisms that produce it will persist and spread.[20] The ongoing Flynn Effect could be seen as *cultural* as opposed to genetic 'assimilate and stretch.'

Let's look at writing systems again. Once some practice such as written accounting is invented and employed by human agents, and is subsequently proven to be of value, then the learning of the practice often finds its way into the domain of formal education. We see examples of this in ancient times. Neugebauer (1935) describes the VAT 8528 tablet, on which is posed the following question:

> If I lent one mina of silver at the rate of 12 shekels (1/60 of a mina) per year, and I received in repayment one talent (60 minas) and 4 minas, how long did the money accumulate?

This is clearly a pedagogic tool with the intention of helping students learn the skill that has been proven to be of value. The processes that led to this type of lesson existing are probably similar to those that drove the revolution in mathematical physics at Cambridge, and the effects of this sort of activity are probably analagous to the effects on IQ of the neurodevelopmental schooling approach to the Flynn Effect.

Genetic assimilation, whether Baldwinian or 'assimilate and stretch', or some other mechanism, modifies the *genome*. In this thesis I will use the terms '*cultome*' and '*teachome*' to designate cultural processes that are both evolvable and that cause the development of cognitive traits. An example of the cultome would be the existence of

[20] If the resulting trait is fitness decreasing and spreads because it is valued, then cultural group selection is likely to select against groups with such values.

number words and examples of the teachome would be particular syllabi in educational systems. Schools become a set of learning outcome facilitators providing a set of learning objects in order that students acquire important cognitive skills. Formal education seems like a reliable inheritance system. It also seems formally modular, and therefore evolvable.

Teachers modify their behaviour to ensure that the required trait is learned. What was previously learned by an individual's adaptive plasticity is moved to the teachome. There is an 'assimilation of past agents' trial and error' process in the construction of education systems, which allows a lineage to 'look' to the future. This allows a more rapid, though less stable (more fragile), cognitive evolution than is possible by genetic assimilation. Learning may indeed find solutions that evolution never would. These innovations can then be further entrenched either genetically via the classical Baldwin effect, or in cultural technologies via teachome effects[21].

I have spent the first half of this chapter introducing the Sapient Paradox and the Neanderthal extinction. I have also outlined three ways in which technological culture can be of relevance not just to developmental outcomes, but also to evolutionary trajectories. These three ways are: cumulative downstream niche construction processes, the evolvabiltiy of technological and pedagogic culture, and a cultural analogue of the 'assimilate and stretch', or Baldwin process of genetic accommodation. With all this in mind, let's move on to explain the two problems of pre-history.

## Explanation 1: The Sapient Paradox

I will now demonstrate how niche construction, the evolvability of pedagogic and symbolic culture, and cultural assimilation processes explain both the sapient paradox and the Neanderthal extinction. The sapient paradox is resolved by understanding the

[21] Elements of culture upon which these sorts of processes act will become cultural analogues of Shea's 'genes-P', which we encountered in chapter four. The culture will persist because of its relationship with the trait being selected for. Hence, the elements of culture acquire teleofunctions and feature in phylogenetic explanations.

effect of an increasingly complex technological and symbolic environment, and accumulating changes to pedagogic methods and content.

Sproull and Kiesler (1992a, 1992b) describe two levels of effects of technologies. First level effects are those that are planned and often efficiencies gained go hand in hand with offsetting consequences. Second level effects result in unpredictable and uncontrolled technology use and are associated with changes in social practice and social contexts. New technology can therefore lead to a new way of thinking by fundamentally changing how people work and interact with others. Something as simple as using ochre for camouflage may have effected this, as it was then co-opted as a signal of group identity. New technologies may have important unexpected cognitive effects. Furthermore, cultural technologies have their developmental effects not only, indeed not even mainly, in the generation that constructs them. Cultural technologies can affect the cognitive development of *the next (and subsequent) generations*. This is cumulative downstream niche construction. None of these resulting cognitive skills appeared at the same time as anatomical moderns. They were necessarily subsequent to the emergence of *Homo sapiens*.

If we think that the first sapiens were not cognitively like moderns, then what happened? One approach is to see the invention and use of physical symbols as a key change. Some have argued that this is some sort of *mark* of fully sophisticated modern cognition. However, it may be a *cause* of such cognition. There exist at least two different kinds of symbolism, 'physical symbols' are not a homogenous category. There are symbols that have a physical function, such as dyeing one's body with ochre, this both camouflages and marks an individual as a member of a group. And then there are purely abstract symbols such as glyphs:

> 'understood one way, the ability of a mind to use and understand symbols really is a signature of cognitive sophistication. But those are not the symbols used in group self-identity, and hence are not the symbols whose presence becomes obvious in the Upper Paleolithic and Late Stone Age.' (Sterelny 2011, p. 63).

As a 'signature of cognitive sophistication', writing emerged only after the social problem space in which it was required existed. Similarly, the appearance of ochre,

beads and other public ‘symbols’, are an effect of demographic and social change. People only need to badge their identity once their social worlds become dense enough. Nevertheless once effected, these features of the developmental environment would almost certainly start driving new forms of cognition as we saw with number words for example.

An example of the prehistoric symbolic storage of important information is the Montgaudier baton (10,000 b.p.). This is an engraved antler segment from inland in southwest France (figure 24).

Figure 24: The Montgaudier Baton

The baton depicts particular spring plants, notably a flower which can be identified. In addition, in this view, a bull and cow seal are depicted, along with a male salmon with the characteristic hooked bottom jaw it develops having begun its spawning run upstream from the Atlantic. The salmon's run coincides with seals congregating on beaches for breeding, both these events occurring in spring. The baton can thus be 'read' as containing the message, 'when these plants appear it is time to journey down river to the sea for good hunting' (Lock & Gers 2011). It is important to note that you cannot simply draw this without considerable practice in acquiring the skills for drawing in the first place. Think how long children must spend engaging with crayon and chalk technologies in order to draw rudimentary drawings let alone realistic representations. Note how the pictures on the baton contribute to enhancing fidelity of cross-generational adaptive information flow, and also act as developmental resources for children learning how to use symbolism. So some

innovation, which was learned by an individual, is now assimilated into a formal structure and can be taught to subsequent agents. These descendents can then stretch the knowledge, and it can be further assimilated into the teachome.

The sapient paradox is the Flynn Effect writ large. In our contemporary world IQ increases may originally have caused increased complexity of popular media, or the increases may be driven by it, but either way the ratchet will begin. I propose that an analogous cognitive, not merely cultural, ratchet got underway with the emergence of moderns 100,000 years ago. New technology emerged due to accumulation (such as ways, and then new and better ways, to make weapons) and larger databases of know-how and facts about the physical and biological world could be preserved such as migration patterns (e.g. the baton above, perhaps cave art as well). But also, as well as quantitative accumulation of cognitive capital, there would have been significant cognitive change. A new suite of possibilities would have opened for *Homo sapiens*. This is almost certainly the case when we remember the new cognitive abilities gained by those lucky enough to develop in a context of number words, or rich social media, or external memory storage devices. And taking into account the features of evolvable systems that I have described above, the Hoyleyness of some forms of culture, the possibility of assimilation of important causes of cognitive development into the cultome and teachome, then this explains recent cognitive evolution and adaptation as well. These models are important and relevant even for us today because they point to our capacity for change (Gamble, 2007, p. 34).

Symbolism does not mark the sudden emergence of cognitive modernity, but it drives cognitive modernity. ‘There is no other technological advance whose effects on human history rival those of this one’ (Nickerson, 2005, p. 25). Nickerson catalogues a host of cognitive amplification devices such as slide rules and memory aids, all of which amplify our capabilities. Sometimes these scaffold cognition, sometimes they form proper parts of cognitive systems. There are good reasons to believe that there was not a burst of cultural and behavioural take off, rather, the gradual, accelerating and distributed appearance of the signs of behavioural modernity in the archaeological record suggests a more diffuse and patchy build up. New privileged

causes of cognitive traits (such as ochre, art, tallys, symbols and teaching methods) emerged piecemeal and with increasing frequency and impact *after* we became anatomically modern.

Trial and error, which has always been important, can result in lucky accidents that become regular events in generation N+1. A classic example of this is potato washing in the macaque monkey (Avital & Jablonka 2000). Just like genetic mutations, cultural mutations can drive evolution. This coupled with high fidelity, high bandwidth inheritance explains human uniqueness and also cultural diversity. Anatomically modern *Homo sapiens* did not emerge pre-equipped with the results of trial and error learning and experimentation. They needed to actually undertake the process. And over time changing and accumulating technological and symbolic innovations changed the developmental context for their children. Once this began to happen new cognition was born.

Traditional views of cognitive evolution were that complex culture was the product of a new mind. And that a new mind was the product of a new genome. But the delay between achieving modern brain size and the arrival of truly complex culture casts doubt on this as sole explanation. Similarly, the glimpses of symbolic thought and linguistic ability in primates suggests that cultural thought may not be uniquely human anyway (Savage-Rumbaugh et al. 1986). Recent authors (Mithen, 2000; Renfrew, 2008) have noted that through material engagement of various sorts (especially with symbols) humans have enhanced their cognitive capacities by forming new extended systems, which constitute cognition (see chapter three). Material culture is not just the product of a new brain, but it is part of a new mind. One causal channel that hasn't been much emphasized in the literature, however, is that material culture and cultural technologies are also the *cause* of a new brain.

As I have explained in chapters one and two, cultural technologies are not only produced by brains, and not only extend brains, but they also alter brains through developmental processes. Renfrew (2008), Mithen (2000), Donald (1991), Sterelny (2011), and others tend to only emphasize two, or two and a half of these processes.

The creation of culture is emphasized, as is the integration of culture and minds (Donald's external memory field). But less is said about the changing developmental context and the altered cognitive development of the next generation. This is certainly consistent with some of these models, but more seems to be said about accumulating epistemic features of culture. As well as being useful in their own right such cultural technologies as games, role-plays, notching to record quantities, story-telling, marking days or seasons, formal teaching, accounting, body adornment for institutional reasons, richer language, painting, and so on are all likely to have effects on the cognitive development of those children exposed to them (consider all the effects of cultural technologies on cognitive development we encountered in chapter one, and the mechanism underlying them discussed in chapter two). In fact the importance of child development for the evolution of cognition probably can't be emphasized enough (see Bjorkland 2007).

Donald's model of the origins of our modern mind is one of punctuated radical change. There are three transitions. First is the emergence of mimesis (the ability to voluntarily retrieve a stored motor memory). This allows any voluntary action to be stopped, replayed, edited, and performed without the presence of contextual cues. The second change in Donald's progression is language. The third transition, and the one I am most focused on in this section, is the 'externalization of memory'. This evolutionary move is a technological innovation, and hence entirely cultural, whereas the previous two transitions were biological. There is well documented evidence of visuo-symbolic culture appearing throughout the Upper Paleolithic. Donald argues that the effects of this externalization of memory would be four-fold. It is in these details that we start to see a hint of the effects of cultural technologies on brain development.

1. There were now radical new properties in collective storage and retrieval systems.
2. There had to be a major redeployment of cerebral resources towards establishing 'literacy' 'modules' in the brain. This has to come at a cost, remember the light patterns shone into chicks' eggs in chapter two.

3. The physiological basis for this reorganization lies in neuronal epigenesis and plasticity (explained in chapter two).
4. The role of biological working memory is forever changed.

Hopefully chapter one illustrated that radical changes could have accumulated in many domains not just memory and literacy. Affect, concepts of number, attention, all sorts of cognitive devices would start to change with radical technological change. Donald's net can be cast wide.

We are context-dependent individuals, and thanks to cumulative niche construction and 'assimilate and stretch' phenomena, we are therefore a moving target of cognitive investigation. This is largely because it is our 'basic human nature to annex, exploit, and incorporate non-biological stuff deep into our mental profiles' (Clark 2003, p. 198). Such non-biological causes of our cognitive profiles include ochre, cave art, symbols, and more recently tutorial notes, mathematical syllabi, social network site layouts, and digital avatars. The Sapient Paradox continues even today. There are many other potential examples, it has been argued that SMS messaging is driving a linguistic renaissance. The explosion in font diversity leads us to interpret words as feelings not just having semantic content. And the emergence of emoticons ☺ is bringing a different sort of symbol, sometimes animated, to our lexicon. These are all individually small changes, but small ongoing changes are the foundation of Lyell's geological explanations. Once cultural technologies are in place those which are highly Hoyley can evolve adaptations. Further efficiencies can be gained in the learning process through Baldwinian and assimilation effects. Cognitive improvements can accumulate without genetic change.

My cultural technological account is not the final word. There are a host of different factors all contributing to explaining the sapient take-off. New learning dynamics and social transitions enable the accumulation of cultural technologies. Sterelny (2011) proposes that the transition to behavioural modernity was a multifactorial process that basically had the effect of increasing fidelity and bandwidth of cross-generational information transfer. His model argues for the importance of longer periods in a

juvenile learning state, the expansion of material culture, changes to the social environment that permitted information pooling and teaching, and further evolution of adaptations for cultural learning. These disparate processes combine to form a feed-forward loop, which once it begins to establish, further elaborates. A key piece in the puzzle for Sterelny is the rise of apprentice learning from a milieu of what was previously mere trial and error learning. I am not denying that such mechanisms have an important role to play. But it is a role that frequently uses technology as leverage in cognitive development.

Those skeptical of this kind of story may object that Neanderthals pose a problem for the theory. Neanderthals probably possessed roughly the same cognitive resources as *Homo sapiens* and were probably endowed with similar degrees of neural plasticity. Why did they die out while humans were using their high bandwidth, high fidelity cultural inheritance and apprentice learning to prosper?

## Explanation 2: The Neanderthal Extinction

The last Neanderthal went extinct approximately 28,000 years ago (Finlayson et al. 2006). There are many hypotheses pertaining to the extinction of the Neanderthals. They can be broadly grouped into competition, hostility, assimilation, and climatic causes. The lack of clear data suggests that assimilation with sapiens is unlikely. However it is not clear whether Neanderthals simply failed to adapt to the dramatic climatic pressures of the last glacial maximum, or were outcompeted by early moderns, perhaps lacking some cognitive or behavioural advantage. In favour of the climatic approach is the argument that even combining sapiens and Neanderthal population numbers the species were below the carrying capacity of the environment (Finlayson 2004). Climatic explanations are also the most parsimonious (Stewart 2006). So let's assume this basic structure is the right way to approach the Neanderthal extinction. The idea here is that whether or not there was some cognitive deficiency Neanderthals possessed when compared to sapiens, that other factors can entirely explain the disappearance of Neanderthals. Climatic or ecological

models seem to apply especially if Neanderthals and sapiens were allopatric, living in non-overlapping geographical areas.

I want to further assume that Neanderthals possessed similar anatomical, neural and genetic resources to *Homo sapiens*. My aim is to show that there can be a cognitive explanation for the Neanderthals' demise, *even if* they are demonstrated to have had similar potential to humans, and *even if* climate plays a significant role. Indeed, Finlayson (2004), Zilhao (2010), and Sterelny (2011) play down any possible cognitive differences between Neanderthals and sapiens.

Sterelny (2011) doesn't want to posit cognitive differences with sapiens as explaining the Neanderthal extinction. He thinks that rather than a cognitive/competitive explanation (e.g. Mithen 2005, O'Connell 2006) there is an ecological/social explanation:

> 'I think it quite likely that [Neanderthals] were pushed over the brink by negative feedback loops between demography and environment. These loops eroded the preconditions of high volume social learning.' (Sterelny 2011, p. 75)

This explanation is both parsimonious and avoids the problem of invisible evidence that cognitive explanations suffer from. In a nutshell the climate changed, forcing Neanderthals to marginal geographical locations. The choice to retreat seemed better because to adapt to a way of life other than ambush hunting with no sexual division of labour would mean crossing a fitness valley that was too risky an option. Existing in new refuges (which still had the carrying capacity to support them) meant the Neanderthals were very vulnerable, this is because smaller group numbers struggled to retain skill bases, specialization, and the redundancy required to transmit culture successfully. The social and information network was disrupted. Fidelity and bandwidth decreased due to ecological pressure and social disruption, and the great adaptation of the Neanderthals (i.e. their ability to transmit culture so effectively) became their biggest problem. They were dependent on culture and technology which was no longer able to be transmitted effectively. Even chance encounters with contemporaries and the possibility of horizontal transmission couldn't help because

the required skills need years of practice and close living to obtain. The ratchet took a dive and eventually fizzled out. In the language of this chapter the Hoyle qualities of Neanderthal culture decayed, and fragile know-how that had been culturally Baldwinized fell apart, no need to posit cognitive differences between Neanderthal and sapiens.

However, I have spent much of this thesis arguing that even when the same basic resources are available, agents' cognition can vary markedly due to contextual factors. It is on this basis that it is unlikely that the cognition of recent (say 10,000 years ago) *Homo sapiens* was the same as that of those alive 100,000 years ago. It also seems unlikely that the general cognition of Neanderthals just prior to their extinction even remotely resembled that which they possessed a few millennia earlier. This is because Neanderthal social structure and technology, and hence their general context of development was in disarray by that point. If significant cognitive differences did not exist between sapiens and Neanderthals at the beginning of the extinction sequence, then they almost certainly did at the end of it. My further claim is that as soon as Neanderthal cognition *began* to become impoverished due to an impoverished social and technological environment then this would have driven a negative feedback loop that further eroded cognitive capability. I will now give further details along these lines.

One question is this, if Neanderthals did have contact with sapiens (as some accounts suggest) and they were an equally as intelligent species, why did Neanderthals not realise full symbolic potential? Neanderthals were probably pre-adapted for symbolism. Indeed they possessed rudimentary symbolic practices. Zilhao et al. (2010) claim to have found Neanderthal shell jewellery, and they probably used ochre as well, although this may just have been as camouflage. Given the right developmental conditions fully abstract symbolic thought could probably have arisen. Neanderthals it seems lagged somewhat behind *Homo sapiens* in the invention of several key technologies. These include representational symbolism (there are no known Neanderthal cave paintings for example), and probably lacked significant division of labour.

Once the concept of abstract representation really took hold in humans then there was a burst of cultural invention. We see the same phenomena with writing systems and mathematical systems. There have been a lot of refinements of the cultural anchors for our mathematical cognition. Chinese, Indian and European systems of math had their different limitations, but the European system has kept evolving at a rapid pace once it was realized that there was control of levels of abstraction (e.g. i for the square root of negative numbers). This sort of thing may explain why Neanderthals may have failed to evolve symbols. Once the process is underway it is an exponential ratchet, but it takes deep insight to get it off the ground. And even if they had lots of contact with *Homo sapiens*, Neanderthals who didn't understand yet what symbols or representations *were* would not be able to grasp their significance.

Had Neanderthal developed symbolic thought a few millennia earlier this may well have been a catalyst to a host of inventions that could have eased the ecological pressure. This could have happened through externalizing memory (perhaps cave art), facilitating calendars (the Montgaudier baton), trade, supporting teaching, and so on. However in order to pick up symbolism (and all which that entails such as cave art, number concepts, and so on) from sapiens, then the right Neanderthals (children) needed to come into extended (months/years) contact with the right sapiens. It takes a long time to learn to draw or use notches for numbers. And it must occur at the right stage of development (as Akbar's deprived children demonstrate). Furthermore, if sapiens were indifferent to the existence of Neanderthals then simple observational learning on the part of Neanderthals would probably not do. For it is difficult to even understand the target of learning if you are not aware of what a symbol is. Indeed, it may simply be that the two species only met quite rarely, as there is little firm evidence of regular interaction (Sterelny 2011). There are many contingencies that may explain why humans currently exist everywhere on the planet and Neanderthals are extinct.

Had it not been for rapid cognitive evolution in part caused by the invention of symbolism and rich language, sapiens may have gone extinct too. The Hoyley nature

of their newfound explicit teaching-apprentice setup and the process of cultural Baldwinization meant that they could track environmental changes in adaptive ways. *Homo sapiens*, I argue, were cognitively just leading the Neanderthals when the climatic problems hit. This was not due to any inherent difference in genes, neurons, anatomy or other such developmental resources, but due to the slight lag in Neanderthal technology and symbolic mastery. This was enough to make all the difference in their ability to adapt. The start of the downward spiral suffered by Neanderthals (which in all likelihood was triggered ecologically) degraded the Hoyley nature of their cultural system, and the pace of change meant that genetic evolution could not produce minds adapted to the new circumstances.

Neanderthals by 50,000 years ago were a 'species under stress' (Sterelny, 2011, p. 75). Ecological changes, including massive glaciation, had driven the population into a scattering of small refuges. Once confined to small populations due to ecological changes, the Neanderthals' social, encultured life-way became a hindrance to survival. Hunting became dangerous and less effective, shorter life-spans led to a shorter adolescence as young ones are needed to hunt, intergroup interactions cease, skills lost are not replenished, and so on. But this is not all. As well as these effects, developmental environments are changing. As group size (and effective group size due to isolation), and technological know-how, diminish we would see the opposite of all the effects described in chapter one. Popular culture, such as fireside stories, may become thin and omit important detail. It may be that the popular culture of Neanderthal was more like 'Dragnet' than 'The Sopranos'. Their games more like throwing stones than taking on fictional identities. It is unclear what the nature of the language of Neanderthals was, but as they experienced demographic decline and isolation what language they had would regress. A smaller population leads to impoverished retention or innovation of vocabulary. If rich language can drive cognitive flexibility perhaps Neanderthal lacked rich language. There may have been a lack of cultural redundancy and duplication of important knowledge bases. Hence there would be less chance of Cambridge tutorial-style experiments, no educational scaffolding, lower IQs, and perhaps affect changes through lack of play. It is easy to

see how all this could have caused extinction in difficult circumstances for a species that depended heavily on cultural inheritance and the cognition that culture caused.

Despite the fact that at one point Neanderthals had sophisticated culture such as burial, brain extraction, ochre, cooking, leopard hide clothes, caring for the sick, beads, and so on, Neanderthals went extinct, because, as argues Sterelny, they regressed in their technological and social complexity. This has been well documented in the case of indigenous Tasmanians who lost the ability to make bone tools and ceased fishing (Diamond 1993). This is likely to have also occurred when moderns first reached Australia, largely due to the dramatic demographic and social structure consequences of the founder effect (where the characteristics of the founding members of a population limit what can descend from them). A specific example of demographic shifts causing technological regress can be seen in Denmark where lithic arrowhead complexity was lost and point shape changed due to a population drop around 5-6000 BCE (Edinborough 2009).

One of the key points surely is that learning things takes a long time. Children devote hours and hours and hours in playing to learn these things. With all their ecological stress, and the demand for 'all hands on deck' so to speak, Neanderthals probably didn't have the required opportunities to play, and especially to play with technologies required to scaffold rich cognitive development. The technologies that children play with determine in part their cognitive abilities. Furthermore, social processes determine the child's play materials (Gauvain 2001). The technological context will determine this too. As technological culture decays due to ecological and then demographic reasons, affordances of child's play diminish. Socially offered options diminish and we see cognitive regress. This drives further regress in the developmental options for the next generation, and suddenly we have *reverse* cumulative niche construction.

Similar potential, but a simple cognitive evolutionary lag when compared to sapiens may explain why Neanderthals didn't adapt and experience a transformation into behavioural modernity in time to combat climatic change.

Basically the explanation of the Neanderthal extinction looks like the nuclear apocalypse scenario we imagined in the introduction. Alongside Sterelny's technological regress arguments, we have to suppose that Neanderthals also suffered cognitive regress, in the exact opposite way to what I have argued for the sapiens' take-off. Neanderthals had the potential to be as intelligent as us, the simply lucked-out on the right developmental contexts. Being slightly behind the mark when climate change hit, their trajectory was very different. Even when culture *is* Hoyley it is a slow, step-wise and cumulative process to evolve complex adaptation. However, the reverse process can be rapid and dramatic.

So the argument is not that *pre-existing* cognitive differences and competition between sapiens and Neanderthals caused extinction. But that cognitive differences *resulting from* an inability to adapt, possibly caused by a slight lag in behavioural, technological and therefore cognitive evolution, occurred after the insult. This meant no coming back for the Neanderthals.

This is the framework I suggest employing in order to explain the actual mechanisms through which climate change and subsequent demographic disruption drove the extinction sequence. There were several contributing causes and as part of the explanation we must look for the relationship between technological culture and cognitive development. With this in mind it would be interesting to search and see if key technologies disappear from the archaeological record and when? This is how one would set about seeking the fine details of an actual lineage explanation in this case.

So the Neanderthal objection to the sapient paradox story fails. We can explain the Neanderthal extinction with similar machinery to that that which explains sapiens' success. Small differences in technological repertoire and small lags in cognitive evolution might have been all that was needed to make or break a species in a certain ecological context.

I am interested in this thesis in technology and cognition. So even if Sterelny's 'ecological explanation' for the Neanderthal extinction is correct in broad strokes, we can still ask did technology (or its lack) help the ratchet on the way down, and were there cognitive effects of this? We can add grain to the explanation by understanding the cognitive impoverishment that might come from losing or no longer acquiring key technological advances. Advances that would act as privileged causes of cognitive development.

The sapient paradox is explained by a cultural-technological-developmental ratchet. The Neanderthal extinction is explained by precisely the reverse effect. Several negative events ratchet off each other and drive a technological and cultural regress. To illustrate, think again of the Cambridge tutorial example I gave in chapter three. Imagine if the tutorials were abolished, writing of notes in class ceased to be encouraged, and paper and pencils were no longer provided. Students would begin to perform poorly in mathematical physics, this would result in a dearth of skilled lecturing staff and further regress of physics education.

## Summary and Conclusions

In the movie 'The Matrix' Neo spends time learning to fight. The process by which he learns is that an interface rod is plugged into his brain through a socket already installed in his skull. The relevant information required to, say, use Kung Fu, is uploaded to his brain in a format that means he can employ the technique immediately. The point is that if there is the right information in the environment, presented to the agent in the right way, and there exist adaptations for uptake of the information, then teaching and learning and the development of novel cognitive devices can be seamless. However, if the information is degraded, perhaps because the relevant culture has lost its Hoyleyness, or if the teachome or cultome are incomplete, or the appropriate pedagogical structures are absent, or the learning window is disrupted, or crucial content fails to get recorded or memorized, then the system generating adaptive developmental environments may break down. Stable, high-fidelity, high-volume flow depends on individual cognitive adaptation, learning

environments, and demographic support (to buffer resources and supply innovation) (Sterelny, 2011). It seems that Neanderthals may have lacked two of these. They were demographically disrupted, and the learning environments were degrading. Mental representations were probably not being preserved across generations and so instructional intention becomes lost. Technologies would be going extinct, and with them developmental environments. Cognition would have been decaying and then there would have been no return.

The important conclusion to draw from this chapter is not necessarily a fine-grained explanation of these two puzzles. I've emphasized that I am not attempting to give an actual sequence of events that explains these problems. I'm merely trying to explain what would constitute such an explanation. Would Ayla, the fictional Cro-Magnon, raised by Neanderthals, have succeeded in using her sapiens smarts to profit? The answer is probably not, especially if her developmental environment was impoverished. Would a Neanderthal, raised by Cro-Magnons, be capable of representational art and fundamental mathematical correspondences? I think the answer is probably. Essentialist thinking must give way to causative, contextualised, developmental processes, and this applies even today. Finally, it will not have escaped the reader's attention that the Neanderthal extinction itself is an effect with myriad complex causes of its development...

## Final Summary and Conclusions

In chapter one, I demonstrated that cultural technologies play a causal role in psychological trait development. In chapter two, I explained the constructivist neural mechanisms underpinning these phenomena. In chapter three, I argued that external media further support cognition and afford us novel and powerful capabilities. In part two, I examined causal roles in developmental systems and argued that these cultural technologies sometimes play an important privileged role in development. This reasoning helped us to explain away the puzzling nature of the sapient paradox and the Neanderthal extinction, in part by reference to changing cultural and technological environments.

My overall point is this: *As we build our worlds, we build our minds*. Culture has a significant role to play in the development of psychological traits. Largely this is through formal education. But cultural technology, too, has a significant role to play in these processes. Cultural technology causes developmental changes in the brain, cultural technology supports psychological functions, and cultural technology partially constitutes some loops of cognitive processing.

In an often underestimated book, Merlin Donald (1991) describes three phases in the evolution of human cognition. I believe that in most respects Donald has got the story exactly right. Our lineage went through important transitions generating mimesis, language and symbol use. However a lot has happened within the period denoted by his 'third phase' (indeed, a lot has happened since Donald wrote in 1991!) and these changes need to be fleshed out. The import of doing this is that Donald's third phase is a period of rapid human cognitive evolutionary change. Prehistory is a long time and the third phase is comparatively brief. This suggests that the changes Donald has described are very much only a part of an ongoing process of human cognitive evolution.

The result is that there are important elements of human cognitive evolutionary theory that are relevant for us today, here and now. I have argued that developmental environments shape human minds to some degree. We have all heard of the 'poor upbringings' that many social misfits have had. We know that 'good' and 'bad' environments will effect child cognitive development. However, if we look at environmental change over time, we see that some developmental contexts can ratchet further change in developmental context and, therefore, in cognition. By tracing the context of child development and how that context changes over time and place, and by studying the effects that this has had on human minds and culture, throughout history and prehistory, we can say something important about human potential and human nature. We can identify what causes human minds to change, and what contextual modifications can do to human thought. What will be the effect on human thought of the emergence of pervasive virtual environments? What would human cognition be like in the aftermath of nuclear war? Were Neanderthals human? If we take distinctively human cognition to be that which is enabled by external symbolic media then perhaps 'humans' are very, very recent indeed. These are the sorts of questions that this thesis should give us the conceptual tools to answer.

Ever since biologically modern humans emerged there has been a dramatic wealth of technological and task-focused knowledge and artifacts produced. This changes the developmental environment dramatically, and the result is ongoing, rapid, cumulative changes to the brain and cognition. The really important changes actually came *after* the Neanderthals. It was only in the last 10,000 years, really only in the last 5,000 that we see the emergence of writing. With the possible exception of language, this is without a doubt the most important transformative technology that humans have invented. Today we experience total immersion in a virtual world of written texts. Written symbols constitute a massive portion of our physical environments. But the ability to write, and the existence of a developmental niche utterly teeming with written symbols are only one factor in an intricate developmental web. We have seen arguments for the role of genes, social structure, pedagogic practices, attentional biases, symbolic technologies, children's games, and a host of other factors intertwining to hone the development of just a single cognitive trait. Given that this

process of cognitive evolution (in part driven by novel technologies) is not only ongoing, but actually accelerating, and given the wealth of technologies that we are currently creating which seem to have an impact on cognitive development, we are going to need some way to tease apart this dire swamp of causes. The time when genetic determinism was taken seriously is long gone. The developmental system is intricate and interrelated. If we have any hope of intervening sensibly on it by setting education policy, or remedial contexts, or in designing technological aids to cognition, then we must be able to understand, in tractable fashion, developmental causation. Genes certainly 'reach' and produce amino acid sequences, cultural technologies 'reach' and effect cognitive concepts, somewhere in the middle the two mix and mingle. Hopefully the tool I have introduced in chapters four through six goes some way toward suggesting how we might untangle this mess and apportion causal responsibility in order to both explain and generalize over these complex systems.

Furthermore, we should expect to see variability in human behaviour and cognition across place and time. Sterelny is probably right to claim that 'behavioural modernity' is a flawed concept and that high-bandwidth and fidelity and an apprentice learning structure are the important innovations. Shea (2011) echoes this sentiment, 'as an analytical construct behavioural modernity is deeply flawed at all epistemological levels' (p. 1). Behavioural and cognitive variability is exactly what we would expect given the role that neuroconstructive processes have been playing ever since we had a 'modern' brain. 'Progression' from being essentially primitive to being essentially modern is not the way to look at things. So much is context-dependent, and so much could vanish overnight if contexts were to change. There is no forward momentum in evolution (Godfrey-Smith 2009). Such is the precarious and fragile nature of human psychological traits.

But things can go the other way too. I have argued for population-wide changes in what we are capable of achieving. The advent of number words is one example of this. So, too, is the general shift in IQ in the twentieth century. We are extremely adept now at offloading static information to our environments, as lists, as notes, as texts, in accounting, in law, in every domain of human endeavour. But technology

now permits us to offload dynamic processes too. We can program computers to think for us, and to keep processing while we turn our attention to other tasks. What will become of human capabilities if basic programming skills, and new ways of representing problems, become a population norm? Human psychology is a moving target of investigation.

# REFERENCES

Alexopoulos, D. (1997) Urban vs rural residence and IQ. *Psychological Reports*, 80, 851-860.

Auel, J. (1980). *Clan of the Cave Bear*. London: Hodder & Stoughton.

Avital, E., and Jablonka, E. (2000). *Animal Traditions: Behavioural Inheritance in Evolution*. Cambridge: Cambridge University Press.

Barkow, J., Cosmides, L., Tooby, J. (1992). *The Adapted Mind: evolutionary psychology and the generation of culture*. Oxford: Oxford University Press.

Barrett, L. (2009). The Future of Psychology: Connecting Mind to Brain. *Perspectives on Psychological Science*, 4(4): 326-339.

Bartlett, C., Vowels, C., Shanteau, J., Crow, J., Miller, T. (2009). The Effect of Violent and Non-violent Computer Games on Cognitive Performance. *Computers in Human Behaviour*, 25: 96-102.

Barton, D., & Hamilton, M. (1996). Social and Cognitive Factors in the Historical Elaboration of Writing. In A. Lock & C. Peters (eds) *Handbook of Human Symbolic Evolution*. Oxford: Clarendon Press.

Bergstrom, C., & Rosvall, M. (2009). The transmission sense of information. *Biology and Philosophy*, 26(2): 159-176.

Bjorklund, D. (2007). *Why Youth is Not Wasted on the Young*. Malden: Blackwell.

Blair, C., Gamson, D., Thorne, S., Baker, D. (2005). Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurobiology of the prefrontal cortex. *Intelligence* 33: 93-106.

Boyd, R., Richerson, P. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Burdick, K., Lencz, T., Funke, B., Finn, C., Szeszko, P., Kane, J., Kucherlapati, R., Malhotra, A. (2006). Genetic variation in DTNBP1 influences general cognitive ability. *Human Molecular Genetics*, 15(10): 1563-1568.

Buss, D. (1995). *The Evolution of Desire: strategies of human mating*. Basic Books.

Capsi, A., Williams, B., Kim-Cohen, J., Craig, I., Milne, B., Poulton, R., Schalkwyk, L., Taylor, A., Werts, H., Moffitt, T. (2007). Moderation of Breastfeeding Effect on the IQ by Genetic Variation in Fatty Acid Metabolism. *Proceedings of the National Academy of Sciences*, 104(47): 18860-65.

Cardinelli, L., Frassinetti, F., Brozzoli, C., Urquizar, C., Roy, A., Farne, A. (2009). Tool-use induces morphological updating of the body schema. *Current Biology*, 19(12): R478-479.

Carpenter, P., Just, M., Shell, P. (1990). What One Intelligence Test Measures: A Theoretical Account of the Processing in the Raven Progressive Matrices Test. *Psychological Review*, 97(3): 404-431.

Carruthers, P. (2006). *The Architecture of the Mind*. Oxford: Oxford University Press.

Cartwright, N., Shomar, T., & Suarez, M. (1995). The Tool Box of Science: Tools for building models with a superconductivity example. In, W. Herfel et al. (eds.) *Theories and Models in Scientific Processes: proceedings of AFOS*. Rodopi Editions, 137-149.

Catrou, F. (1826). *History of the Mogul Dynasty in India from its foundation by Tamerlane in the year 1399 to the accession of Aurengzebe, in the year 1657* . London: J.M. Richardson.

Cavelli-Sforza, L., & Feldman, M. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press.

Cha, J., & Young, A. (2000). Huntington's Disease. In, Bloom & Kupfer (eds.) *Psychopharmacology: the fourth generation of progress*. American College of Neuropsychopharmacology.

Chan, P., & Rabinowitz, T. (2006). A cross-sectional analysis of videogames and attention deficit hyperactivity disorder symptoms in adolescents. *Annals of General Psychiatry*, 5(16). http://www.biomedcentral.com/content/pdf/1744-859X-5-16.pdf

Changizi, M., & Shimojo, S. (2005). Character Complexity and redundancy in writing systems over human history. *Proc. R. Soc. B*. 272: 267-275.

Clark, A. (1997). *Being There*. Cambridge MA.: MIT Press.

Clark, A. (2003). *Natural Born Cyborgs*. Oxford: Oxford University Press.

Clark, A. (2006). Author's Response to Symposium on Natural Born Cyborgs. *Metascience*. http://www.cogs.indiana.edu/andy/metasciencereply.pdf

Clark, A. (2008). *Supersizing the Mind*. Oxford: Oxford University Press.

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis* 58: 10-23.

Confer, J., Easton, J., Fleischman, D., Goetz, C., Lewis, D., Perilloux, C., Buss, D. (2010). Evolutionary Psychology: Controversies, questions, prospects, and limitations. *American Psychologist*, 65(2):110-126.

Cooper, J. (2004). Babylonian Beginnings: the origin of the cuneiform writing system in comparative perspective. In, S. Houston (ed.) *The First Writing: Script Invention as History and Process*. Cambridge: Cambridge University Press.

Cooper, R., & Zubeck, J. (1958). Effects of enriched and restricted early environments on the learning of bright and dull rats. *Canadian Journal of Psychology*, 12(3): 159-164.

Cowie, F. (1998). *What's Within? Nativism Reconsidered*. Oxford: Oxford University Press.

Daley, T., Whaley, S., Sigman, M., Espinosa, M., & Neumann, C. (2003). IQ on the rise, the Flynn effect in rural Kenyan children. *Psychological Science,* 14(3): 215-219.

Davies, P. (2009). Some evolutionary model or other: Aspirations and evidence in evolutionary psychology. *Philosophical Psychology*, 22(1): 83-97.

De Cruz, H. (2008). An Extended Mind Perspective on Natural Number Representation. *Philosophical Psychology*, 21(4): 475-490.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56: 384-398.

Dennett, D. (1995). *Darwin's Dangerous Idea*. Simon & Schuster.

Dennett, D. (2000). Making Tools for Thinking. In D. Sperber (ed.) *Metarepresentation*. Oxford: Oxford University Press.

Diamond, J. (1993). 10,000 years of Solitude. *DISCOVER Magazine*. March Issue.

Dickens, W., & Flynn, J. (2001). Heritability Estimates Versus Large Environmental Effects: The IQ Paradox Resolved. *Psychological Review*, 108(2): 346-369.

Diringer, D. (1968). *The Alphabet: a key to the history of mankind*. New York: Funk & Wagnalls.

Donald, M. (1991). *The Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge Mass. Harvard University Press.

Dye, M., Green, C., & Bavelier. (2009). The Development of Attention Skills in Action Video Game Players. *Neuropsychologia* 47: 1780-1789.

Edinborough, (2009). Population History and the Evolution of Mesolithic Arrowhead Technology in South Scandinanvia. In, S. Shennan (ed.) *Pattern and Process in Cultural Evolution*. Berkeley: University of California Press.

Eisenberg, N. (2002). Empathy-related emotional responses, altruism, and their socialization. In R. J. Davidson & A. Harrington (Eds.). *Visions of compassion: Western scientists and Tibetan Buddhists examine human nature* (pp. 131-164). London: Oxford University Press.

Everett, D. (2005) Cultural constraints on grammar and cognition in Piraha. *Current Anthropology*, 46: 621-646.

Feigenson, L., Dehaene, S., Spelke, E. (2004). Core Systems of Number. *TRENDS in Cognitive Sciences*, 8(7): 307-314.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an Action Video Game Reduces Gender Differences in Spatial Cognition. *Psychological Science*, 18(10): 850-855.

Finlayson, C. (2004). *Neanderthals and modern humans. An ecological and evolutionary perspective*. Cambridge: Cambridge University Press.

Finlayson, C., Pacheco, F., Rodriguez-Vidal, J., et al. (2006). Late Survival of Neanderthals at the Southernmost Extreme of Europe. *Nature*, 443: 850-853.

Finlayson, C. (2009). *The Humans who went extinct: Why Neanderthals died out and we survived*. Oxford: Oxford University Press.

Flynn, J. (1987). Massive IQ gains in 14 Nations: What IQ tests really measure. *Psychological Bulletin*, 101: 171-191.

Flynn, J. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, 54: 5-20.

Flynn, J. (2007). *What is Intelligence?* Cambridge: Cambridge University Press.

Fodor, J. (1983). *The Modularity of Mind*. Cambridge MA.: MIT Press.

Frank, M., Everett, D., Fedorenko, E., Gibson, E. (2008). Number as a Cognitive Technology: Evidence from Piraha Language and Cognition. *Cognition* 108: 819-824.

Gamble, C. (2007). *Origins and Revolutions: human identity in earliest prehistory*. Cambridge: Cambridge University Press.

Gauvain, M. (2001). *The Social Context of Cognitive Development*. New York: Guilford.

Gee, J. (2008). Good Videogames, the Human Mind, and Good Learning. In: T. Willoughby & E. Wood (eds.) *Children's Learning in a Digital World*. Malden: Blackwell.

Gelb, I. (1963). *A Study of Writing*. Chicago: University of Chicago Press.

Giere, R. (1988). Explaining Science. Chicago: University of Chicago Press.

Gilbert, S. (2002). Genetic determinism: The battle between scientific data and the social image in contemporary molecular biology. In: A. Grunwald, M. Gutmann, & E. Neumann-Held (eds.), *On human nature: Anthropological, biological, and philosophical foundations* (p. 121-141). Berlin: Springer.

Gilbert, S. (2003). Evo-Devo, Devo-Evo, and Devgen-Popgen. *Biology and Philosophy*, 18: 347-352.

Gleick, J. (1992). *Genius: The Life and Science of Richard Feynman*. New York: Vintage Books.

Godfrey-Smith, P. (1999). Genes and Codes. In V. Gray Hardcastle (ed). *Where Biology Meets Psychology*. Cambridge Ma.: MIT Press.

Godfrey-Smith, P. (2000). On the Theoretical Role of 'Genetic Coding'. *Philosophy of Science*, 67: 26-44.

Godfrey-Smith, P. (2001). On the Status and Explanatory Structure of Developmental Systems Theory. In, S. Oyama, P. Griffiths, R. Gray, (eds.), *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge MA.: MIT Press.

Godfrey-Smith, P. (2006a). Information in Biology. In M. Ruse, Hull, D. (eds.) *The Cambridge Companion to the Philosophy of Biology*. Cambridge: Cambridge University Press.

Godfrey-Smith, P. (2006b). Theories and Models in Metaphysics. *The Harvard Review of Philosophy*, 14(1): 4-19.

Godfrey-Smith, P. (2009). *Darwinian Populations and Natural Selection*. Oxford: Oxford University Press.

Goldstein, P. (2010). The Age of Inception: Are you too old to get it? *The Dominion Post*, Saturday 7 August.

Goody, J. (1977). *The Domestication of the Savage Mind*. Cambridge: Cambridge University Press.

Goody, J. (1987). *The Interface Between the Written and the Oral*. Cambridge: Cambridge University Press.

Goody, J. & Watt, I. (1963). The consequences of literacy. *Comp. Stud. Soc. Hist*. 5, 304-45.

Gordon, P. (2004). Numerical Cognition Without Words: Evidence from Amazonia. *Science*, 306: 496-499.

Gottlieb, G. (1971). *Development of Species Identification in Birds*. Chicago: Chicago University Press.

Gould, S. (2002). *The Structure of Evolutionary Theory*. Chicago: Cambridge Ma.: Harvard University Press.

Gould, S., & Lewontin, R. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A critique of the adaptationist programme. *Proc. R. Soc. Lond*. B, 205: 581-598.

Grice, H. (1957). Meaning. *Philosophical Review*, 66: 377-388.

Griffiths, P., & Gray, R. (1994). Developmental Systems and Evolutionary Explanation. *Journal of Philosophy*, 91: 277-304.

Griffiths , P., & Gray, R. (2005). Three ways to misunderstand developmental systems theory. *Biology and Philosophy*, 20: 417-425.

Grifftiths, P., & Stotz, K. (2000). How the mind grows: a developmental perspective on the biology of cognition. *Synthese,* 122(1-2): 29-51.

Halverson, J. (1992). Goody and the implosion of the Literacy Thesis. *Man,* 27(2): 301-317.

Harlow, H., & Suomi, S. (1971). Social Recovery by Isolation-Reared Monkeys. *Proceedings of the National Academy of Sciences*, 68(7): 1534-1538.

Harrell, M. (2008). No computer program required: even pencil and paper argument mapping improves critical thinking skills. Carnegie Mellon University Research Showcase. Department of Philosophy. http://repository.cmu.edu/philosophy/350

Hashimoto, R., Noguchi, H., Hori, H., Ohi, K., Yasuda, Y., Takeda, M., Kunugi, H. (2009). Association between the dysbindin gene (DTNBP1) and cognitive functions in Japanese subjects. *Psychiatry and Clinical Neurosciences*, 63: 550-556.

Hauser, M., Carey, S., Hauser, L. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys*. Proceedings of the Royal Society of London*. B, 267: 829-833.

Hauser, M., Tsao, F., Garcia, P., Spelke, E. (2003). Evolutionary foundations of number: spontaneous representation of numerical magnitudes by cotton-top tamarins. *Proc. R. Soc. Lond*, 270: 1441-1446.

Hobson, J., Pace-Shott, E., & Stickgold, R. (2000). Dreaming and the Brain: Toward a cognitive neuroscience of conscious states. *Behavioural and Brain Sciences*, 23: 793-1121.

Houston, S. (ed.) (2004). *The First Writing: Script Invention as History and Process*. Cambridge: Cambridge University Press.

Iriki, A., Tanaka, M., Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque psotcentral neurons. *Neuroreport,* 7(14): 2325-2330.

Jablonka, E. (2002). Information: Its interpretation, Its Inheritance, and Its Sharing. *Philosophy of Science*, 69(4): 578-605

Jablonka, E., & Lamb, M. (2005a). *Evolution in Four Dimensions: genetic, epigenetic, behavioural and symbolic variation in the history of life*. Cambridge MA.: MIT Press.

Jablonka, E., & Lamb, M. (2005b). The evolution of information in the major transitions. *Journal of Theoretical Biology*, 239: 236-246.

Jackson, F., & Pettit, P. (1992). In Defense of Explanatory Ecumenism. *Economics and Philosophy*, 8(1): 1-21.

Johnson, S. (2005). *Everything Bad is Good for You: How today's popular culture is actually making us smarter*. New York: Rimerhead Books.

Johnson, S., Hennessy, E., Smith, R., Trikic, R., Wolke, D., Marlow, N. (2009). Academic attainment and special educational needs in extremely preterm children at 11 years of age: the EPICure Study. *Arch. Dis. Child. Fetal Neonatal Ed*. Online 12 March: http://wrap.warwick.ac.uk/566/1/WRAP_Wolke_Epicure.pdf

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice.

Karmiloff-Smith, A. (2009). Nativism vs. Neuroconstructivism: Rethinking the Study of Developmental Disorders. *Developmental Psychology*, 45(1): 56-63.

Kelleher, C. & Pausch, R. (2005). Stencils-based tutorials: design and evaluation. Conference on Human Factors in Computing Systems, Portland, Oregan. p. 541-550.

Kelleher, C., Pausch, R., Kiesler, S. (2007). Storytelling Alice Motivates Middle School Girls to Learn Computer Programming. CHI Proceedings, San Jose, CA.

Kitcher, P. (2001). Battling the Undead: How (and How Not) to Resist Genetic Determinism. In R. Singh et al. (Eds) *Thinking About Evolution, Historical, Philosophical and Political Perspectives*. Cambridge: Cambridge University Press. 396-414.

Konrath, S., O'Brien, E., Tsing, C. (2011). Changes in Dispositional Empathy in American College Students Over Time: A Meta-Analysis. Pers. Soc. Psychol. Rev., 15(2): 180-198.

Laland, K. (2008). Exploring gene-culture interactions: insights from handedness, sexual selection and niche construction case studies. *Phil. Trans. R. Soc. B*. 363: 3577-3589.

Laland, K., Odling-Smee, J., Myles, S. (2010). How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Reviews – Genetics*, 11: 137-148.

Levy, N. (2004). Evolutionary Psychology, Human Universals, and the Standard Social Science Model. *Biology and Philosophy*, 19: 459-472.

Lewis, D. (2000). Causation as Influence. *Journal of Philosophy*, 97: 182-97.

Lewis, M. (2004). Environmental Complexity and Central Nervous System Development and Function. *Mental Retardation and Developmental Disabilities Research Reviews* 10: 91-95.

Lewontin, R. (1995). *Human Diversity*. New York: Scientific American Library.

Lickliter, R. (1990). Enhanced prenatal auditory experience facilitates species-specific visual responsiveness in bobwhite quail chicks (*colinus virginianus*). *Journal of Comparative Psychology*, 105(1): 89-94.

Linton, R. (1933). *The Tanala, A hill tribe of Madagascar*. Chicago: Field Museum of Natural History.

Lipton, J., & Spelke, E. (2004). Discrimination of Large and Small Numerosities by Human Infants. *Infancy*, 5(3): 271-290.

Lock, A. & Gers, M. (2011). The Cultural Evolution of Written Language and its Effects: A Darwinian Process from Prehistory to the Modern Day. In E. Grigorenko, E. Mambrino, D. Preiss (eds.) *Writing: A Mosaic of New Perspectives.* Psychology Press.

Lock, A. & Peters, C. (eds.) (1996) *Handbook of Human Symbolic Evolution*. Oxford: Clarendon Press.

Locke, J. & Bogin, B. (2006). Language and Life History: A new perspective on the development and evolution of human language. *Behavioural and Brain Sciences*, 29: 259-325.

Lombard, L. (1992). Causes and Enablers: A Reply to Mackie. *Philosophical Studies*, 65(3):319-322.

Lorenz, K. (1937). The Companion in the Bird's World. *The Auk*, 54(3): 245-273.

Lundborg, G. (2000). Brain Plasticity and Hand Surgey: an Overview. *The Journal of Hand Surgery European Edition*, 25(3): 242-252.

Lyell, C. (1830) [1997]. *Principles of Geology, Being an Attempt to Explain the Former Changes in the Earth's Surface by Reference to Causes Now in Operation*. London: Penguin.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1): 1-25.

Maguire, E., Gadian, D., Johnsrude, I., et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, 97: 4398-403.

Maguire, E., Woollett, K., Spiers, H. (2006). London Taxi Drivers and Bus Drivers: a structural MRI and neuropsychological analysis. *Hippocampus*, 16: 1091-101.

Mameli, M., & Bateson, P. (2006). Innateness and the Sciences. *Biology and Philosophy*, 21(2): 155-188.

Mareschal, D., Johnson, M., Sirois, S., Spratling, M., Thomas, M., Westermann, G. (eds.) (2007). *Neuroconstructivism Vol. One and Vol. Two*. Oxford: Oxford University Press.

Marr, D. (1982). *Vision*. W.H. Freeman & Co.

Marshack, A. (1972). *The Roots of Civilization*. New York: McGraw-Hill.

Masterson, M (1970). The Nature of a Paradigm. In, I. Lakatos and A. Musgrave (eds.) *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.

Maynard Smith, J. (2000). The Concept of Information in Biology. *Philosophy of Science*, 67: 177-194.

McLuhan, M. (1962). *The Gutenberg Galaxy: The making of typographic man*. University of Toronto Press.

Menary, R. (2007). Writing as Thinking. *Language Sciences,* 29: 621-632.

Mercier, H. (2006). Some ideas to study the evolution of mathematics. *Language and Culture*, 39(3): 351-377.

Mesoudi, A., Whiten, A., Laland, K. (2006). Towards a unified science of cultural evolution. *Behavioural and Brain Sciences*, 29: 329-383.

Millikan, R. (1984). *White Queen Psychology and Other Essays for Alice*. Cambridge MA.: MIT Press.

Mioduser, D., Tur-Kaspa, H., Leitner, I. (2000). The Learning Value of Computer-Based Instruction in Early Reading Skills. *Journal of Computer Assisted Learning*, 16: 54-63.

Mithen, S. (2000). Mind, Brain and Material Culture: an archaeological perspective. In P. Carruthers, A. Chamberlain (eds.) *Evolution and the Human Mind: Modularity, Language and Meta-cognition*. Cambridge: Cambridge University Press.

Mithen, S. (2003). *After the Ice*. London: Weidenfeld & Nicolson.

Mithen, S. (2005). *The Singing Neanderthals: The origins of music language mind and body*. London: Weidenfeld and Nicolson.

Neugebauer, O. (1935). *Mathematische Keilschrift-Texfe*. Berlin: Springer-Verlag.

Nickerson, R. (2005). Technology and Cognitive Amplification. In R. Sternberg & D Preiss (eds.). *Intelligence and Technology: The impact of tools on the nature and development of human abilities*. Mahwah, NJ: Lawrence Erlbaum Associates. pp. 3-27.

Nisbett, R. (2003). *The Geography of Thought: How Asians and Westerners Think Differently, and Why*. New York: Free Press.

O'Connell, J. (2006). How Did Modern Humans Displace Neanderthals: Insights from hunter-gatherer ethnography and archaeology. In, N. Conrad (ed.) *When Neanderthals and Modern Humans Met*. Tubingen: Kerns. 43-64.

Odenbaugh, J. (2006). Models. Manuscript, http://legacy.lclark.edu/~jay/Biologymodels.pdf

Odling-Smee, J., Laland, K., Feldman, M. (2003). *Niche Construction: the neglected process in evolution*. Princeton NJ: Princeton University Press.

Olson, D. (2005). Technology and Intelligence in a Literate Society In R. Sternberg & D Preiss (eds.) *Intelligence and Technology: The impact of tools on the nature and development of human abilities*. Mahwah, NJ: Lawrence Erlbaum Associates. pp. 55-67.

Oyama, S. (1985). *The Ontogeny of Information*. Cambridge: Cambridge University Press.

Oyama, S. (2000). *The Ontogeny of Information* (2nd Ed). Duke University Press.

Oyama S., Griffiths, P., Gray, R. (2001). *Cycles of Contingency: Developmental Systems and Evolution*. MIT Press.

Parasuraman, R., & Greenwood, P. (2007). Individual differences in attention and working memory: A molecular genetic approach. In, A. Kramer, D. Wiegmann, & A Kirlik (eds.) *Attention: From theory to practice* (pp. 59-72). New York: Oxford University Press.

Paulsen, J., Langbehn, D., Stout, J., Aylward, E., Ross, C., Nance, M., Guttman, M., Johnson, S., MacDonald, M., Beglinger, L., Duff, K., Kayson, E., Biglan, K., Shoulson, I., Oakes, D., Hayden, M. (2008). Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *Journal of Neurology, Neurosurgery and Psychiatry*, 79(8): 874-880.

Pearl, J. (2000): *Causality*. New York: Cambridge University Press.

Pennington, B., McGrath, L., Rosenburg, J., Barnard, H., Smith, S., Willcutt, E., Friend, A., DeFries, J., Olson, R. (2009). Gene x Environment Interactions in Reading Disability and Attention-Deficit/Hyperactivity Disorder. *Developmental Psychology*, 45(1): 77-89.

Perez-Montoro, M. (2007). *The Phenomenon of Information*. Maryland: Scarecrow Press.

Piaget, J. (1955). *The Child's Construction of Reality.* London: Routledge.

Pica, P., Lemer, C., Izard, V., Dehaene, S. (2004). Exact and Approximate arithmetic in an Amazonian Indigene group. *Science,* 306: 499-503.

Pigliucci, M., Murren, C., Schlichting, C. (2006). Review: Phenotypic plasticity and evolution by genetic assimilation. *The Journal of Experimental Biology*, 209: 2362-2367.

Pinker, S. (1997). *How The Mind Works*. London: Penguin.

Potochnik, A. (2007). Optimality Modeling and Explanatory Generality. *Philosophy of Science*, 74(5): 680-691.

Putnam, H. (1975). The Nature of Mental States. In: *Mind, Language and Reality.* Cambridge: Cambridge University Press.

Putnam, H. (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

Qin, Y., Carter, C., Silk, E., Stenger, V., Fissell, K., Goode, A., Anderson, J. (2004). The change of the brain activation patterns as children learn algebra equation solving. *Proceedings of the National Academy of Sciences*, 101(15): 5686-5691.

Quartz, S., Sejnowski, T. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioural and Brain Sciences*, 20: 537-596.

Ramachandran (1999). *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. Quill.

Rambusch, J., Susi, T., Ziemke, T. (2004). Artefacts as Mediators of Distributed Social Cognition: A Case Study. http://www.his.se/PageFiles/5675/cogsci04.rambusch.etal.pdf?epslanguage=sv

Renfrew, C. (2008). *Prehistory: the Making of the Human Mind*. New York: Random House.

Richerson, P., & Boyd, R. (2005). Not By Genes Alone. Chicago: University of Chicago Press.

Ridley, Matt. (2003). *Nature via Nurture*. London: Fourth Estate.

Rosenberg, A. (2006). *Darwinian Reductionism*. Chicago: University of Chicago Press.

Russo, F. (2009). *Causality and Causal Modelling in the Social Sciences: Measuring Variations*. Springer.

Russo, F. (2010). Are Causal Analysis and Systems Analysis Compatible Approaches? *International Studies in Philosophy of Science*, 24(1): 67-90.

Savage-Rumbaugh, E. Sue. (1986). *Ape Language: From conditioned response to symbol.* New York: Colombia University Press.

Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction*, 19(6): 481-94.

Schwenkreis, P., Tom, S., Ragert, P. et al. (2007). Assessment of sensorimotor cortical representation asymmetries and motor skills in violin players. *European Journal of Neuroscience*, 26: 3291-3302.

Searle, J. (1980). Minds, Brains and Programs. *Behavioural and Brain Sciences*, 3: 417-457.

Shannon, C. (1948). A mathematical theory of communication. *Bell Syst Tech J*, 27:379-423.

Shea, J. (2011). Homo sapiens is as Homo sapiens was: Behavioural variability versus behavioural modernity in Paleolithic archaeology. *Current Anthropology*, 52(1): 1-35.

Shea, N. (2006). Representation in the genome and in other inheritance systems. *Biology and Philosophy*, 22: 313-331.

Shennan, S. (ed.) (2009). *Pattern and Process in Cultural Evolution*. Berkeley: University of California Press.

Shu, W., Lu, M., Zhang, Y., Tucker, P., Zhou, D., Morrisey, E. (2007). Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development*, 134: 1991-2000.

Shultz, T., Mysore, S., Quartz, S. (2007). Why Let Networks Grow? In D. Mareschal, M. Johnson, S. Sirois, M. Spratling, M. Thomas, G. Westerman (eds.) *Neuroconstructivism Vol. Two*. Oxford: Oxford University Press.

Skelton, C. (2008). Methods of Using Phylogenetic Systematics to Reconstruct the History of the Linear B Script. *Archaeometry*, 50(1): 158-176.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, 15: 201–293.

Sperber, D. (2000). Metarepresentations in an Evolutionary Perspective. In, D. Sperber (ed.), *Metarepresentations: A Mulitdisciplinary Perspective*. Oxford: Oxford University Press.

Sproull, L., & Kiesler, S. (1992a). Secondary Effects of Technology. *Computerworld*, Editorial.

Sproull, L., & Kiesler, S. (1992b). Group decision making and communication technology. *Organizational Behavioural and Human Decision Processes*, 52(1): 96-123.

Steen, R. (2009). *Human Intelligence and Medical Illness: Assessing the Flynn Effect*. Springer.

Stephens, B., & Vohr, B. (2009). Neurodevelopmental outcome in the premature infant. *The Pediatric Clinics of North America*, 56(3): 631-646.

Sterelny, K. (2001). Niche Construction, Developmental Systems, and the Extended Replicator. In, S. Oyama, P. Griffiths, R. Gray (eds.) *Systems and Cycles of Contingency: Developmental Evolution*. Cambridge MA.: MIT Press.

Sterelny, K. (2003). *Thought in a Hostile World*. Malden: Blackwell.

Sterelny, K. (2004). A Review of Evolution and Learning: The Baldwin Effect Reconsidered, edited by Bruce Weber and David Depew. *Education and Development*, 6(4): 295-300.

Sterelny, K. (2010). Minds: Extended or Scaffolded? *Phenomenology and the Cognitive Sciences: Special Issue 4E Cognition: Embodied, Embedded, Enacted, Extended*, 9(4): 465-481.

Sterelny, K. (2011). *The Evolved Apprentice*. Cambridge Ma.: MIT Press.

Sterelny, K., & Godfrey-Smith, P. (2007). Biological Information. *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/information-biological/

Sterelny, K., Griffiths, P. (1999). *Sex and Death*. Chicago, Chicago University Press.

Sterelny, K., Smith, K., Dickison, M. (1996). The Extended Replicator. *Biology and Philosophy*, 11(3). 377-403.

Sternberg, R., & Preiss, D. (eds.) (2005). *Intelligence and Technology: The impact of tools on the nature and development of human abilities.* Mahwah, NJ: Lawrence Erlbaum Associates.

Stewart, J. (2006). Review of, C. Finlayson: *Neanderthals and Modern Humans. An ecological and evolutionary perspective*. *Journal of Quaternary Science*, 21(2): 206-210.

Stiles, J. (2009). On Genes, Brains, and Behaviours: Why Should Developmental Psychologists Care About Brain Development? *Child Development Perspectives*, 3(3): 196-202.

Stotz, K. (2010). Human nature and cognitive-developmental niche construction. *Phenomenology and the Cognitive Sciences*. *Special Issue 4E Cognition: Embodied, Embedded, Enacted, Extended*, 9(4): 483-501.

Sutton, J., Harris, C., Keil, P., Barnier, A. (2010). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the Cognitive Sciences*. *Special Issue 4E Cognition: Embodied, Embedded, Enacted, Extended*, 9(4): 521-560.

Swing, E., & Anderson, C. (2008). How and What do Videogames Teach? In: T Willoughby & E Wood (eds.) *Children's Learning in a Digital World*. Malden: Blackwell.

Taglialatela, J., Russell, J., Schaeffer, J., Hopkins, W. (2009). Visualizing Vocal Perception in the Chimpanzee Brain. *Cerebral Cortex*, 19(5): 1151-57.

Taglialetela, J., Russell, J., Schaeffer, J., Hopkins, D. (2008). Communicative Signaling Activates 'Broca's' Homolog in Chimpanzees. *Current Biology* 18: 343-348.

Tahiroglu, A., Celik, G., Avci, A., Seydaoglu, G., Uzel, M., Altunbas, H. (2009). Short-Term Effects of Playing Computer Games on Attention. *Journal of Attention Disorders*. 13(6): 668-676.

Thompson, R., Oden, D., Boysen, S. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behaviour Processes*, 23: 31-43.

Thornton, A., & McAuliffe, K. (2006). Teaching in Wild Meerkats. *Science*, 313: 227-229.

Toga, A., & Thompson, P. (2005). Genetics of Brain Structure and Intelligence. *Annual Review of Neuroscience*, 28: 1-23.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, Mass.: Harvard University Press.

Tomasello, (2008). *Origins of Human Communication*. MIT Press.

Trigger, B. (1968). *Beyond History: the methods of prehistory*. Holt, Reinhart & Winston.

Trigger, B. (2004). Writing Systems: a Case-study in Cultural Evolution. In, S. Houston (ed.) *The First Writing: Script Invention as History and Process*. Cambridge: Cambridge University Press.

Turner, B. (2007). Defining an Epigenetic Code. *Nature Cell Biology*, 9(1): 2-6.

Turner, C., Lewis, M., King, M. (2003). Environmental Enrichment: Effects on Stereotyped Behaviour and Dendritic Morphology. *Dev Psychobiol,* 43: 20-27.

Twardy, C. (2004). Argument Maps Improve Critical Thinking. *Teaching Philosophy*, 27, 95-116.

van Gelder, T. (2001). How to improve critical thinking using educational technology. In G. Kennedy, M. Keppell, C. McNaught, & T. Petrovic (eds.), *Meeting at the crossroads: proceedings of the 18th annual conference of the Australian Society for computers in learning in tertiary education* (pp. 539-548). Melbourne.

Vasalou, A., Joinson, A., & Pitt, J. (2007). Constructing my online self: Avatars that increase self-focused attention. Paper presented at the conference on human factors in computing systems. (pp. 445-448) San Jose, USA.

Vlassis, J. (2008). The Role of Mathematical Symbols in the Development of Number Conceptualization: The Case of the Minus Sign. *Philosophical Psychology*, 21(4): 555-570.

Vygotsky, L. (1962). *Thought and Language*. Cambridge, MA: The MIT Press.

Vygotsky, L. (1978). *Mind in Society: The development of higher psychological processes*. Cambridge, MA.: Harvard University Press.

Walker, S., Wachs, T., Meeks Gardner J., Lozoff, B., Wasserman, G., Pollitt, E., Carter, J., & the International Child Development Steering Group. (2007). Child Development: risk factors for adverse outcomes in developing countries. *The Lancet*, 369: 145-157.

Warwick, A. (2003). *Masters of Theory: Cambridge and the Rise of Mathematical Physics*. Chicago: Chicago University Press.

Watson, J., Strayer, D. (2010). Supertaskers: Profiles in Extraodinary Multi-tasking Ability. *Psychonomic Bulletin and Review*, 17(4): 479-85.

Weisglas-Kuperus, N., Hille, E., Duivenvoorden, H., Finken, M., Wit, J., van Buuren, S., van Goudoever, J., Verloove-Vanhorick, S. (2009). Intelligence of very pre-term or very low birthweight infants in young adulthood. *Arch Dis Child Fetal Neonatal Ed*, 94(3): F196-200.

West-Eberhard, M. (2003). *Developmental Plasticity and Evolution*. Oxford: Oxford University Press.

Wheeler, M. (2003). Do Genes Code for Traits? In, A. Rojszczak, J. Cachro, G. Kurczewski (eds.) *Philosophical Dimensions of Logic and Science: Selected Contributed Papers*. Dordrecht: Kluwer.

Wheeler, M. (2010). In Defense of Extended Functionalism. In, R. Menary (ed.), *The Extended Mind*. (p. 245-170). Cambridge MA.: MIT Press.

Williamson, S., Hubisz, M., Clark, A., Payseur, B., Bustamante, C., Nielsen, R. (2007). Localizing Recent Adaptive Evolution in the Human Genome. *PLoS Genetics* 3(6): 1-15.

Wilson, M. (2010). The re-tooled mind: how culture re-engineers cognition. *Social Cognitive and Affective Neuroscience*. Online Jan 12th.
http://people.ucsc.edu/~mlwilson/publications/RetooledMind.pdf

Wimsatt, W. & Griesemer, J. (2007). Reproducing Entrenchments to Scaffold Culture. In, R. Sansom & R. Brandon (eds.) *Integrating Evolution and Development: From theory to practice*. MIT Press.

Woodward, J. (2003): *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, J. (2010). Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation. *Biology and Philosophy*, 25(3): 287-318.

Woodward, J. & Hitchcock, C. (2003). Explanatory Generalizations, Part I: A Counterfactual Account. *Nôus*, 37: 1–24.

Yee, N., & Bailenson, J. (2009). The Difference Between Being and Seeing: The Relative Contribution of Self-Perception and Priming to Behavioural Changes via Digital Self-Representation. *Media Psychology*, 12(2): 195-209.

Zhang, J., Burdick, K., Lencz, T., Malhotra, A. (2010). Meta-analysis of genetic variation in DTNBP1 and general cognitive ability. *Biological Psychiatry*, 68(12): 1126-33.

Zhao, Z., Zuber, J., Diaz-Flores, E., Lintault, L., Kogan, S., Shannon, K., Lowe, S. (2010). p53 loss promotes acute myeloid leukemia by enabling aberrant self-renewal. *Genes and Development*, 24: 1389-1402.

Zilhao, J., Angelucci, D., Badal-Garcia, E., et al. (2010). Symbolic use of marine shells and mineral pigments by Iberian Neanderthals. *Proceedings of the National Academy of Sciences*, 107(3): 1023-1028.

Zinkstok, J., de Wilde, O., van Amelsvoort, T., Tanck, M., Baas, F., Linszen, D. (2007). Association between the DTNBP1 gene and intelligence: a case-control study in young patients with schizophrenia and related disorders and unaffected siblings. *Behaviour and Brain Function*, 3: 19. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1864987/pdf/1744-9081-3-19.pdf