Sociobiology and its Application to Environmental Ethics

Jack Hamilton James

Student No. 220057823

A thesis submitted for
the degree of
Doctor of Philosophy
of
the University of New England

April 2017

Acknowledgements

I would like to acknowledge dissertation supervisor, Dr. Tony Lynch. When one marks out a goal of significant challenge that is years in the making there are encountered along the way moments of self-doubt, frustration, and indecision. I can truly say this thesis would not have arrived in Tony's absence.

Furthermore, as a philosophical mentor, I could not have asked for more than what Tony willingly provided. Tony's mastery of content has been readily available throughout my thesis and the source of constant inspiration and amazement. Our online pursuit beyond this thesis, *philosopher.io*, also stands as a testament to Tony's character and open-mindedness to embrace new ideas for delivering philosophical content. With Tony's guidance this journey, although at times difficult, has been an absolute pleasure, and while for philosophers 'absolute' is a term that demands extremely careful use – well, here it is more than warranted.

I would also like to take this opportunity to acknowledge Emeritus Professor Peter Forrest, whose continued encouragement, along with Tony's, during my Master's degree allowed me the confidence to entertain the journey now nearing completion.

Finally, I would like to thank my wife Fiona. As Socrates noted, to agree to be a wife of a philosopher is no easy undertaking, for many reasons that begin with our ugliness and end with our continued argument – so thank you for your love and the patience it grants.

Abstract

This thesis develops a multi-attribute philosophical methodology that allows for the careful but pragmatic use of causal sociobiological theories to further the effective application of environmental ethics in decision making and policy construction frameworks. In doing this it considers sociobiological reductionism in the context of a commitment to cultural evolution and cognitive neuroscience.

Certification

Statement of Authors' Contribution Proforma & Statement of Originality:

I certify that the substance of this thesis has not already been submitted for any degree and is not currently being submitted for any other degree or qualification.

I certify that this thesis has been edited by Elite Editing Services and Tony Lynch strictly in accordance with sections D and E of the Australian Standards for Editing Practice.

I certify that any all help received in preparing this thesis and all sources used have been acknowledged in this thesis.

Signature: Date:

18/4/17

Jack Hamilton James
Student No. 220057823

Contents

List of Tables

List of Figures

List of Abbreviations

ACC	anterior cingulate cortex
CBA	cost benefit analysis
DLPFC	dorsolateral prefrontal cortex
DPT	dual process theory
DT	decision theory
EEE	embedded environmental ethic
EEG	electroencephalography
EFEC	event feature emotion complex
EUT	expected utility theory
fMRI	functional magnetic resonance imaging
GAT	guided activation theory
HR	human resources
MCDA	multi-criteria decision analysis
MEG	magnetoencephalography
MMT	modular myopia theory
OFC	orbitofrontal cortex
PCC	posterior cingulate cortex
PET	positron emission tomography
PP	precautionary principle
SD	sustainable development
SEC	structured event complex
SEU	subjective expected utility
SMH	somatic marker hypothesis
SPECT	single photon emission computed tomography
SRR	social response reversal
STS	superior temporal sulcus
TPJ	temporoparietal junction
UCD	universal causal determinism
VMPFC	ventromedial prefrontal cortex
WCED	World Commission on Economic Development

Chapter 1: Introduction

1.1 Placement of this Research in the Existing Body of Literature

Environmental ethics can be broken down into several categories of enquiry that often intersect, including:

1. The origins and history of environmental ethics.
2. Environmental ethics and politics.
3. Environmental ethics and religion.
4. The question of intrinsic value.
5. Anthropocentrism and environmental ethics.
6. The application of ethical theories in response to the current environmental crisis.
7. The future of environmental ethics.

The ethical theories considered in these categories can be further divided into three sub-categories (with many variations): *consequentialism* (John Stuart Mill, Jeremy Bentham), *deontology* (Immanuel Kant) and *virtue ethics* (Aristotle).

The present research is primarily concerned with the sixth category, *the application of ethical theories in response to the current environmental crisis*, and it does so in a generally *consequentialist* way. The ultimate practical concern is with ecological preservation and sustainable living, where this concern is embedded in a *naturalist* account of the origin and functionality of human decision-making. The idea is to explore the biological and cultural causal mechanics on which supervene our thinking and behaviours as they impact on the environment and, having done this, to use this knowledge to shape or 'nudge' us in a properly sustainable direction.

1.2 Thesis Concept

1.2.1 Short

To do this, this research explores recent developments in cognitive neuroscience, the evolution of morality, evolutionary psychology, and cultural evolution.[1]

I argue in favour of the importance and relevance of such naturalist accounts, though only if approached through a tempered, careful philosophical methodology that pays attention to the spaces between facts and values in any integrative proposition. I endorse Bernard William's point that science informs what we can do, and hence, by Kant's principle, what we ought to do. After analysing the relevant sociobiological theory (Chapter 3 through 9) using the defined methodology (Chapter 2), I seek to demonstrate the value of such research in the field of decision theory (DT) and environmental policy construction (Chapter 10).

Sociobiological theory has historically had a poor relationship with ethics, but we can no longer ignore the relevance of emerging sociobiological research. The current environmental crisis grows worse every day, but here is an opportunity to build a bridge upon a tightrope; a metaphor that represents both the distance between two towering but distinct fields (science and ethics), and the path humanity now treads as we seek (to the extent we do) to prevent environmental collapse and its destructive civilizational impacts.

1.2.2 Extended

> *Moral science is not something with a separate province. It is physical, biological, and historic knowledge placed in a humane context where it will illuminate and guide the activities of men.*[2]

> *A normative theory that is uninformed as to the workings of the brain, or is impossible to implement in a brain (or machine), will most likely not be useful for making our world better.*[3]

[1] For a succinct explanation of potentially relevant environmental ethics, see Table A-1 in Appendix A.
[2] J. Dewey, *Human Nature and Conduct: An Introduction to Social Psychology*, in O. Flanagan, H. Sarkissian, and D. Wong, 'Naturalizing Ethics', in *Moral Psychology*, vol. 1, *The Evolution of Morality: Adaptions and Innateness*, ed. W. Sinnott-Armstrong (Cambridge: Massachusetts Institute of Technology Press, 2008), 17.
[3] Gerd Gigerenzer, 'Moral Intuition = Fast and Frugal Heuristics?' in Sinnott-Armstrong, *Moral Psychology*, 2:6.

A meaningful environmental ethics rests on two factual feet. First, and most obviously, facts about the destructive impact of human activities on the environmental foundations of our civilization. And, second, facts about human nature and its placement in nature generally. Normatively it rests on a generalised or collective self-interest pitched at the level of civilizational continuity and a concern for species survivability. We will see that this normativity is itself, in part, intelligibly derivable from the underlying natural facts.

Such facts do not mean we will successfully *change* our behaviours. This is because they bear on what it is we are *capable* of achieving. As philosophers we know, thanks to Moore and Hume, that you cannot brutely derive an *ought* from an *is*, but this is not what we are doing here. We are asking rather for an understanding of those facts that frame or constrain or shape the limits of practical possibility, of what can and what cannot be done, embraced, pursued, etc. 'There seems to be a problem of presenting the ethos [environmental oughts] as a viable position in such domains of action, that is to say a position which links "ought" to "can" in a reasonable way'.[4] Such a 'reasonable way' would seem to demand here, where we are talking of the human impact on the biosphere, that we attend closely both to 'external' facts and processes, and to those distinctively human 'internal' facts of our human nature that are causally significant in producing that external impact and which, therefore, are central to intelligent efforts at changing our behaviours to avert those impacts.

From a simple view, one could argue that it is clear what ethically correct or right behaviour amounts to when it comes to our treatment of the environment: the preservation of the environment and the creation of a sustainable nexus between humanity and earth's natural ecosystems (this, we might think, is a simple, if vital, banality). At a minimum this banality would mean achieving a form of *sustainable development* (SD), which, as articulated by the World Commission on Environment and Development (WCED) in 1987, from the Brundtland report on *Our Common Future*, 'aims to meet the needs of the present without compromising the ability of future generations to meet their own needs'.[5] However, this definition of SD has been criticised as being too vague[6] or even – because it is

[4] Huib Pellikaan and Robert J. van der Veen, *Environmental Dilemmas and Policy Design* (Cambridge, UK: Cambridge University Press, 2002), 231.
[5] Stephen Polasky et al., 'Decision-Making under Great Uncertainty: Environmental Management in an Era of Global Change', *Trends in Ecology and Evolution* 26, no. 8 (2011): 398.
[6] Shunsuke Mangagi, *Technological Change and Environmental Policy: A Study of Depletion in the Oil and Gas Industry* (Cheltenham, UK: Edward Elgar, 2007), 1.

supposed to involve the simultaneous maximisation of three distinct values (ecological, moral and economic) as a neoliberal cover.[7]

Criticisms of SD focus particularly on the language and description of the enterprise, which is, in its 'triple bottom line', win/win/win formulation, likely self-defeating and removed from how we actually live. But no-one, not even critics like Lynch, deny for a moment that the *ambition* bound into SD is a mistake. The destination (SD) is not under question – it is, as we have seen, a banality. The trouble is in making sense of this banality, in a way that opens the way to intelligent planning and effective implementation strategies. I certainly do not discourage a deeper normative appreciation of sustainability, which can add value to the practical realm in the traditional sense. However, it is only once we start to understand the factual, causal linkages—the underlying mechanics of human decision-making—that we can begin to make headway towards effectively realising this crucially human banality.

In response to this idea, Lynch might argue that his attack on SD is not just a matter of criticising its language, but rather of its practicality as typically formulated. Thus:

1. As a 'Guiding Principle' (like the Categorical Imperative or The Principle of Utility), SD is nonsense—for it folds in three distinct functions.
2. This means in practice SD either i) Has one function as strongly dominant, or ii) Involves complex and largely individually conceived trade-offs.
3. In 'real' practice what happens is the former—Strong Dominance—and in the service of 'economic development'.

Lynch's critique of SD amounts to a challenge to the practicality of our acting in accord with my simple banality, for in effect he is asking: Could we 'fix' all this (specify and arrive at SD) from a set of universally agreed 'factual' premises? His answer is No, for however many facts you have, they do not amount to a practical determination for action, and are typically selected on the basis of a prior and suppressed evaluative foundation. Despite all this, I argue that such facts, properly specified and understood, *are* useful for the specification of SD and for its attainment.

[7] Tony Lynch, 'Understanding What Sustainable Development is and What It is Not', *Philosopher.io* (blog), 6 October 2015, http://philosopher.io/Understanding-What-Sustainable-Development-Is-And-What-It-Is-Not.

The first step is to tie this banality to rational self-interest understood, as a matter of the rational self-interest of beings that are fellow members of a species that has developed in and through a natural process (evolution), and whom, in that process, came to be cultural animals, and who, all things being equal, value that capacity.[8] I would add to this that meeting the simple banality is likely best understood or conceived as a matter of how we can do this without *gambling* too much – we do not want to meet our risks by pursuing even riskier, less likely and more dangerous, strategies. Proposals of this risker type come in form of the call for *Radical Revolution* and *Global Value Change*. I think it reasonable to assume we would do better, all things considered, if we avoided such 'utopian' manoeuvres.

The simple banality, and indeed the import of the type of facts given gravity by this thesis, can be situated in a greater scientific recognition; the emergence of the era of the *Anthropocene*. David Grinspoon, a planetary scientist, argues in this book *Earth in Human Hands*, that what differentiates this age from any prior to it is our *awareness* of the *fact* that we are actively shaping the planet. Grinspoon points out that in planetary history life has altered the global climate (for instance, Photosynthetic bacteria), but what differs now is human awareness of the transformative impact of our actions on the biosphere:

> What makes the Anthropocene unprecedented and fully worthy of the name is our growing knowledge of what we are doing to this world. Self-conscious global change is a completely new phenomenon. It puts us humans into a category all our own and is, I believe, the best criterion for the real start of the era. The Anthropocene begins when we start to realise that it has begun…We are the species that can change the world and come to see what we're doing.[9]

Clearly, the simple banality can be situated as a response to this recognition of our "terra-forming" impact and capacity insofar as it is leading – and we *know* it is leading – to processes and outcomes that threaten to undermine the conditions of civilisational continuity, even survival, and are certainly and disastrously impacting on the ecological

[8] Tony Lynch, 'Insuring the Future' *Environmental Values* 10, no.4 (2001) 507-521
[9] David Grinspoon, 'Welcome to Terra Sapiens', *Aeon (Online Magazine)*, 20 December 2016, https://aeon.co/essays/enter-the-sapiezoic-a-new-aeon-of-self-aware-global-change

diversity and robustness of the biosphere itself. In this context, with this awareness, and assuming that general sanity (one might think it deeply embedded in our natures through natural history) which advises us against disaster and warns us against catastrophic risks. we face the challenge of understanding how we have come to be in the position we are in.

The Anthropocene, in Grinspoon's sense of the age in which we know that what we are doing emerges from a scientifically embodied *naturalism*. It is a naturalism that has two sides. That of nature and our impacts on it, and that concerned with understanding why we act in the ways we do and (so) how we might go about altering ourselves and our lives so that we can avoid what we know to be catastrophic outcomes.

In short, we need not merely a naturalistic understanding of such things as the hydrological cycle, climate physics, ecology, biology, soil science and so on; we need a naturalistic understanding of *ourselves*, one that helps us *understand* what *drives* us, and (equally an empirical so natural fact) what it is, in the world in which are driven, that *we value*. We need to know how these things might come together to that we might avoid that general disaster we all want to avoid but can see we are engaged in bringing on.

An Anthropocene naturalism demands we know ourselves as the natural beings we are in that natural world in which we have arisen, live and die. This means looking at the *science* of why we think and behave in ways we do as natural beings; something not straightforwardly or intuitively available to conscious summation. We may be *aware*, but if we do not understand the natural foundations of this awareness, how can we possibly create policies and change that might reliably be thought to bring about sustainability?

This thesis explores the challenges involved in meeting the challenges of the banality through the reductive human sciences as give us insight into understanding why we make the decisions we do and behave as we do. When we ask what actions are best suited to achieve sustainability, a simple reductionist answer is: *those actions most informed by our knowledge of the underlying causal biology of human beings.*

This point is not unappreciated, but has not been directly explored. For the truth is we need a precise meta-paradigm that incorporates ecology and evolutionary facts into an ethical framework:

> The need is increasingly emerging for a connecting epistemological framework able to express a common or convergent tendency of thought and practice aimed at building, among other things, an environmental policy management respectful of the planet's biodiversity and its evolutionary potential.[10]

There is growing recognition in the environmental literature that there is little point in developing a suite of environmental policies based on a putative environmental ethic if that ethic does not fit with (or into) the nature of the creatures that are supposed to adopt and implement it. Tony Lynch and David Wells suggest this when they say:

> It is better to recognize the necessarily anthropocentric nature of human evaluations, and to work on this basis, than to try to promote an ethic among humans which denies themselves.[11]

Owen Flanagan has suggested the same with his principle of *minimal psychological realism*:

> Make sure when constructing a moral theory or projecting a moral ideal that the character, decision processing, and behaviour prescribed are possible, or are perceived to be possible, for creatures like us.[12]

This minimal psychological realism does not mean denying that humans have the capacity for the *intrinsic* valuing, it means only that such a capacity has its natural history and foundations, and so is made and shaped by those matters of fact.

Still, it seems that much ethical theory, and much environmental ethics, has a tendency to spend most of its time thinking and arguing about what has such value "in itself", ignoring the crucial question of how we might effectively realise these values in deliberation, choice and action. The latter has, because of this general lack of interest, has fallen largely into the hands of economists in the form of expected utility theory (EUT) and cost benefit analysis

[10] D. Bergandi, ed., *The Structural Links between Ecology, Evolution and Ethics: The Virtuous Epistemic Circle* (New York: Springer, 2013), 1.
[11] Tony Lynch and David Wells, *The Political Ecologist* (Ashgate, 2000), 15.
[12] Owen Flanagan, *Varieties of Moral Personality* (Cambridge, MA: Harvard University Press, 1991), 32.

(CBA).[13] Environmental ethicists, however, have not staked out and promoted a suitably applicable tool kit, but spent much time attacking the values presupposed by such approaches. Although this has shed much light on biases and limitations of the values so presupposed, it has not birthed any serious alternative to them. Counter principles of application and implementation like the precautionary principle (PP) seek to stem the tide but do not take into account the nature of those who are supposed to effectively wield that principle.

It may be that ethicists have resisted exploring seriously and in depth what it means to take human nature into account when promoting a particular ethical program because to do so is to cross into historically dangerous territory: How can a philosopher consider behavioural causation in terms of substantive biological facts when *resistance* to reductionism (e.g. of the selfish gene kind) has here been typical? Pragmatic considerations of practicality and implementation in a fully naturalistic frame bring into the equation deep philosophical issues concerning causation and determinism which must be addressed in ways that avoid serious contradiction. This is a terribly difficult but unavoidable task, and one that must be faced up to, not merely resisted.

To illuminate the path between the specification of an environmental ethic and its realisation, I explore a specific kind of naturalist account of the kind Joshua Greene explores. As he says:

> First, we must understand the structure of modern moral problems [relating to the environment] and how they differ from the problems that our brains evolved to solve. Second we must understand the structure of our moral brains and how different kinds of thinking are suited to solving different kinds of problems [with regard to the environment].[14]

Meeting these challenges lets us ask of a particular environmental ethic whether it is *executable*. We must be able to execute (in the sense contained in our banality) an effective environmental ethics. It certainly reasonable to think, when engaged in this project, that

[13] Sven Ove Hansson, 'The Harmful Influence of Decision Theory on Ethics', *Ethical Theory and Moral Practice* 13 (2010): 586.
[14] Joshua Greene, *Moral Tribes: Emotion, Reason, and the Gap between Us and Them* (New York, NY: The Penguin Press, 2013), 14.

those whose reflections are informed by the facts relating to the origin and functionality of the human brain are best placed to further this end. The opportunity is there to understand our brains in a naturalistic way – for instance, to understand how we make decisions, often independent of consciousness – and to use this knowledge in reflection of the kind that marks out the Anthropocene itself, to avert disaster and insure sustainability. The scope of possibilities here and the effectiveness of implementation strategies is itself determined by objective fact about the origin and functional capacities of human beings. If, as Lynch insists, trade-offs and (so) the relative weighting of various values (ecological, economic, social) are unavoidable in the pursuit of sustainability, then a naturalistic consideration of those doing the weighting can be expected to bring greater clarity to the choices being made. This should better allow for 'proper balances amongst competing social values'.[15]

'Executable' is a relatively new and, for some, a rather uncomfortable word, for it literally means 'able to be run by a computer'. With this meaning in mind it might seem than an 'executable ethic' would be an 'ethics able to be run by a computer'. Yet 'executable' is also built on a historic term—*execute*—which means 'to put in effect' or 'to carry out' – a usage the long predates computers and any metaphor they may suggest for our brain, or for the metaphysical nature of causation itself. Whilst this language has an individual connotation, it is at the level of the individual with which this thesis is concerned

As this very fact shows, human beings are not *literally* computers, but they certainly are biological entities subject to the natural laws of biology, chemistry and physics. As such, and as recognised by evolutionary psychologists, human beings are composed of information and dedicated processing modules put together over vast periods of time. In this sense we may fairly say that our brains 'compute' information. It may even be that the brain is an *evolved* set of *biological* algorithms.

If a computer cannot execute a software program if that software written without any concern for the computer's hardware constraints, might it not also be the case that a human cannot execute an environmental ethic developed without concern for our evolutionary realized 'hardware'?

[15] Bryan G. Norton, *Sustainability: A Philosophy of Adaptive System Management* (Chicago: University of Chicago Press, 2005).

Information concerning the ways our brains work is potentially a rich and informative source for understanding how we might better organise our conceptions, deliberations and choices so as to ensure environmental sustainability and avoid environmental crisis. In particular we can draw on this information to understand and assess both the destructive behaviours that have let us to our environmental predicament, and the kinds of behaviour open to us to escape this predicament.

This raises the immediate question of those key causal inputs that drive human behaviour.

When it comes to thinking about how we should go about applying an ethic, Robert Traer proposes a linear process of ethical consideration and application:

1. Construct an ethical presumption,
2. Consider the consequences,
3. Make your decision.[16]

The 'ethical presumption' in Step 1 is simply an ethical claim (i.e. that X is good/bad/just/unjust, or that X has an intrinsic value). The 'consequences' in Step 2 are a consideration of the likely outcomes should the presumption be applied to reality, and Step 3 involves weighing Steps 1 and 2 to reach a conclusion**.**

In terms of 'executable ethics' what is missing from this analysis is any concern or investigation into 'the complete factual reality of natural human beings today, as shaped through their evolution'. And that is to say, what is missing is the nature, the set of natural facts, of the agent who is called upon to evaluate, consider and decide.

Such facts are always relevant, regardless of the general ethical stance (consequentialist, deontological, virtue) proposed. But such facts are especially relevant when it comes to our banality, and the need to meet our civilizational threatening environment challenges. On the most fundamental level the task is consequentialist.

[16] Robert Traer, *Doing Environmental Ethics* (West View Press, 2011), 163.

This is because consequentialists are concerned with *consequences*: their job is not merely to define what end state is morally preferable (something our banality defines anyway, and one would hope for all any ethicists, whatever their favoured approach), but to think about how we might best, or at all, achieve that end state. The consequentialist justification of ethics amounts to rule utilitarianism.

Such a consequentialist needs to consider whether a given deontic logic is likely to be widely adopted and used to inform the multitude of real-world decisions that shape our social world. When outcomes matter, human nature matters.[17]

Let us suppose, for example, that we determine that the most effective ethical theory for protecting animal welfare is a deontological assigning of rights rather than a general appeal to benevolence or a call to respect their intrinsic value, because we discover the human brain is more likely to adopt/respect a rights-based approach based on its evolved functioning abilities. Surely this information, given our concern for animal welfare, would lead us to favour the rights-based approach.

Consider too what seems plausibly to be a general fact about human nature; something which has been called *the principle of least disruption.* That principle:

> designates our propensity to avoid discomfort by denying the need for policies and actions which produce unsettling and difficult demands for the active alteration of circumstances otherwise felt to be quite acceptable.[18]

Any serious applicative technique must take into account this natural element of human energy conservation.

The long term environmental philosopher William Ophuls says this our natural constraints in dealing with environmental crisis:

---

[17] Leda Cosmides and John Tooby, 'Can a General Deontic Logic Capture the Facts of Human Moral Reasoning? How the Mind Interprets Social Exchange Rules and Detects Cheaters', in Sinnott-Armstrong, *Moral Psychology*, 1:117.

[18] Tony Lynch and Bert Jenkins, 'Global Warming, Contemporary Politics & the Principle of Least Disruption', *Australian Quarterly*, 79, No.4, 2007.

> We are constrained to see and comprehend the world in certain ways, we therefore struggle to cope intellectually with complex social and environmental conditions that we have created, and we are prone to manias, delusions, and *idea fixes* that may be driven by underlying emotions but that are also the consequences of defects in our thinking process. The gap between what human beings are cognitively capable of and the prevailing conditions in complex mass societies is therefore enormous.[19]

How in such circumstances do environmental ethicists bridge the gap between the quality of their ethics and their effective application? This is one of the most important questions of our time, but only a handful of researchers have attempted to address it.

I address it in this thesis. I aim to bridge environmental ethics (which is concerned with how we ought to treat the environment); ecology, (which is concerned with facts about nature and its interrelations); in the context of the ultimate and pressing practical imperative that is our civilizational protecting banality. This means understanding our evolved nature as the biological organisms (*homo sapiens*) we are because of, and through, our interactions with the broader natural (including here as natural to us, the *cultural*) world. This understanding is essential to us in and as the Anthropocene if we are to have a decent chance at ethical ecology.

This section has attempted to show that while there are issues with the incorporation of scientific facts into the realm of ethics, as evidenced by the history of misuse to be found in things like Social Darwinism, at the same time our environmental crisis is a general one, is very real and imminent, so people want (our banality) that environmental ethics be effective in practice. To be so effective we must (as Socrates insisted) *know ourselves*, for it is we who must act as the beings we are in the world we are in.

This thesis argues that a properly formulated naturalistic reductionism has an important part to play in removing this ignorance. While many insist, not without reason, that oughts can never be derived from the brutely factual, we go wrong if we insist on this division all the way down – down that is, to the banality of not wanting to destroy the conditions for

[19] William Ophuls, *Plato's Revenge: Politics in the Age of Ecology* (Cambridge, MA: MIT Press, 2011), 81.

our species existence. To insist on this level on the *is-ought* gap is to give up on to exit the world of humanity.

1.3 Significance and Target Audience of Thesis

> *Any attempt to transform the massive machinery of modern industrial states into something resembling an ecological benign way of meeting our needs not only involves a continuing and drawn out political and social struggle, but a huge effort in research and innovation.*[20]

By 2030, the world's population will have risen from seven billion people to approximately 8.2 billion. Such predictions demand we develop a navigable route to a sustainable future. Establishing the facts concerning our environmental impacts on the biosphere and so on the foundations of our civilisation is vital, but such truths alone are, as we can see, not enough to motivate people, let alone enough people, to change their behaviour.[21] If *is* and *ought* come together in our banality, that banality is not self-implementing. We must implement it, and to do that we need to know our nature in nature. In the facts of nature, and those of human nature, lie the resources for sustainability.

If, in the end, everyone has a stake in sustainability, not all are primed to actively pursue that goal. Obviously the crucial target audience for this thesis are those persons with already well-reasoned environmental ethics, who by the nature of their position in life have the opportunity to influence policy creation, and who would like to see the most effectively executable realization of sustainability. Given this, the aim (in Elinor Ostrom's terminology) is to positively influence the 'design features' of such policy efforts, utilising the relevant sociobiological facts, and deploying a clear and careful methodology.

The general framework – both because of the nature of our problem ('nature striking back'), and our own nature – lies in my 'commitment to a unified method of naturalism'.

[20] Lynch and Wells, *Political Ecologist*, 61.
[21] Pellikaan and van der Veen, *Environmental Dilemmas and Policy Design*, 4.

Chapter 2: Methodological Validity

2.1 Utopianism

> *Phronesis* is not simply knowledge; it is the capacity and the disposition to put knowledge into practice. The *Phronesis* exercises rather than simply comprehends the virtuous life while securing rather than simply identifying its worldly requirements. *Phronesis* melds practice with principle, stimulating one to act virtuously in concrete situations by integrating and coordinating the various parts of a good life into a well-balanced whole.[22]

How should one understand this 'capacity and disposition to put knowledge into practice' in the service of 'the good life'? I argue that naturalism, and, properly delineated and deployed, a generally reductive naturalism, has a key place in an ethical *phronesis* concerned with sustainability. Understanding our evolved biology helps us understand what it is that makes and shapes us as practical agents, and will help us see what needs to be done and can be done to shift or nudge that biological reality in ways that give us a decent chance of meeting the challenge of our Anthropocene banality.

This ethical *phronesis* of sustainability is not, and could not be, a *utopian* project; where, as Michael Schofield writes, 'to describe a set of ideas as utopian… carries with it the implication that they are impossible to realise in practice.' [23] Sustainability *would* be utopian if we had reason to believe it impossible to realise in practice. To believe that would mean that reflective awareness had no causal powers of its own. But to believe that would be to see such a capacity as, mysteriously, a capacity without effect, and in seeing it that way, to remove it from the influence and forces of natural selection.

Our project is not utopian, but practical. It looks at a human nature in which reflection has its place, and seeks to understand that nature in terms of our natural history. It seeks to

[22] Paul L. Thiele, *Indra's Net and the Midas Touch: Living Sustainably in a Connected World* (Cambridge, MA: MIT Press, 2011), 69.
[23] Michael Schofield, *Plato: Political Philosophy* (New York, NY: Oxford University Press, 2006), 234.

utilise such information so as to show how what, in our pessimistic moments, we might think is 'impossible to realise in practice' (sustainability) is in fact 'practice realisable'.

2.2 Reductionism

2.2.1 Simpliciter

> *I believe that consequentialist and deontological views of philosophy are not so much philosophical inventions as they are philosophical manifestations of two dissociable psychological patterns… Many things have underlying structures that are responsible for making things appear and behave as they do, for giving them their functional properties. And because things have underlying structures, it is possible to refer to something, even make a up definition for it, without really understanding what it is.*[24]

Alongside analogy and metaphor, *reduction* is perhaps the most useful tool for a casual understanding of complex phenomena, and when the object of enquiry is ourselves, it immediately presses on us the question of what 'level' is most conducive to explaining our behaviour—is it the level of quarks, atoms, neurons, modules, networks, math, or that of our full blown humanistic stories? One response is to ask if this is actually a problem for reductionism as such, for would not formal reductionism reveal *all* interacting levels of description? Clearly it matters how one defines reductionism, or what kind of reductionist one is. Nor, of course, is everyone of a reductionist bent. For instance, complexity theory contends that *systems* produce causal outcomes that cannot be derived from an understanding of its parts (the emergent hypothesis).[25] And even for dedicated reductionists there are deep issues centred on how we are to express the results of reductionist analysis. How conceptually formal should neuroscientists be when they make causal claims about the origin of our behaviours? Is Greene on the right track to simply say our ethical theories are just names for causal processes within the brain, which his psychological studies are bringing to light?

These questions have as yet no clear answers, and in many cases we cannot falsify either the reductionist or emergent hypothesis. We are often unable empirically to decide (due to quantum theoretic limits on observation and chaos theoretic limitations on calculation),

[24] Joshua Greene, 'The Secret Joke of Kant's Soul', in Sinnott-Armstrong, *Moral Psychology,* 3:38.
[25] Gregoire Nicolis and Catherine Rouvas-Nicolis, 'Complex Systems', *Scholarpedia* 2, no. 11 (2007): 1473, www.scholarpedia.org/article/complex_systems.

whether the higher-level phenomena we observe actually follow the lower level-laws we have confirmed in simpler circumstances. So, in many cases, the question of reduction vs emergence is unscientific (by Popper's criterion) and mute (because of contextual considerations). Still, what I wish to do here to salvage a reductionism that can be useful for the practical objective of civilizational sustainability in the Anthropocene.

Reductionism as a philosophical method makes this assumption about the nature of complex systems: 'a complex system is nothing but the sum of its parts. An account of it can be reduced to accounts of individual constituents'.[26] The complex systems of concern here are biological systems.

Ontological reductionism is the idea that each particular biological system (e.g. an organism) is constituted by nothing but molecules and their interactions. In metaphysics, this idea is often called *physicalism* (or *materialism*). *Methodological reductionism* is the idea that biological systems are most fruitfully investigated at the lowest possible level, focusing on the decomposition of a complex system into parts. *Epistemic reductionism* is the idea that the knowledge about one scientific domain (typically about higher-level processes) can be reduced to another body of scientific knowledge (typically concerning a lower and more fundamental level).

This thesis is centrally concerned with deploying a methodological reductionism in the context of a naturalistic understanding of our ethical and deliberative capacities. It has, therefore, a deep connection with ontological reductionism, but not to an assumed or *a priori* physicalism, which would rule out the power and impact of cultural and social inputs.

In a very simple reductive sense, such that most think it need not even be mentioned, human ethics in origin and functionality are the product of biological human beings interacting with each other and the world around them. To understand this origin and this functionality one might initially – as we can imagine an alien race studying the human race – study human biology. This makes sense on reductive level, but on a *conceptual* level, it might not. For it seems plausible that such aliens could not predict all our possible

[26] J. Polkinghorne, 'Reductionism', in *The Interdisciplinary Encyclopedia of Religion and Science*, 2002, http://inters.org/reductionism.

behaviours in specific contexts because of the effect of culture or personal experience on (and in) human biology, and in particular, the biology of the brain. Still, one might think that even if such influences have a causal role, that alteration occurs at some biological level below that of the action itself; this is precisely how others argue the whole can act back on the causal parts. However it remains true that this type of reduction has many issues.

The core issue for many critics is that current biological science is ill equipped to deliver a deterministic account,[27] and especially in the case of neuroscience:[28]

> When it comes to behaviour such as displays of cooperation, as opposed to the eye-blink reflex, appealing to innateness is often minimally informative. That is because what mediates the behaviour is neural circuitry, and neural circuitry, as we have seen is the outcome of a gene-gene, gene-neuron-environment, neuron-neuron, and brain-environment interactions…. The basic lesson then is that working backwards from the existence of a certain behaviour to a brain region that supports that behaviour, to the innateness of a function, is, especially in the animals that are prodigious learners, a project fraught with evidential hazard.[29]

Even so, it seems to me that such reductionism is rational *starting point*, in just the same way as one may *start* an explanation of the mind/body problem with the recognition that if one loses one's head (decapitation), one's experience ends. We might think of this as simply an expression of naturalism in so far as it is (minimally and basically) opposed to any kind of supernaturalism.

With this minimal naturalism in place, it is natural to approach the problem of this thesis in terms of *evolutionary theory* and *the brain*. If it turns out to be the case (as we may discover on our journey) that (say) the role of cultural evolution on behaviour has a greater impact or significance than innate biological functionality governed through genes, then we can document the exceptions and understand the relations. With each layer exposed the

[27] E. Mayr, *Toward a New Philosophy of Biology* (Cambridge, MA: Harvard University Press, 1988) 24-38.
[28] E. Racine and J. Illes, *Is Neuroethics the Heir of the Ethics of Genomics?* (Calgary, Alberta: Canadian Bioethics Society, 2004), 1-21.
[29] Patricia S. Churchland, *Brain Trust: What Neuroscience Tells Us about Morality* (Princeton, NJ: Princeton University Press, 2011), 109.

system as a whole is better understood, the forms of the parts are revealed, and potentially practical information can be decoded. A useful and informative example of such a reductionist approach is to be found when we look at the eye. (Figure 2.1).[30]

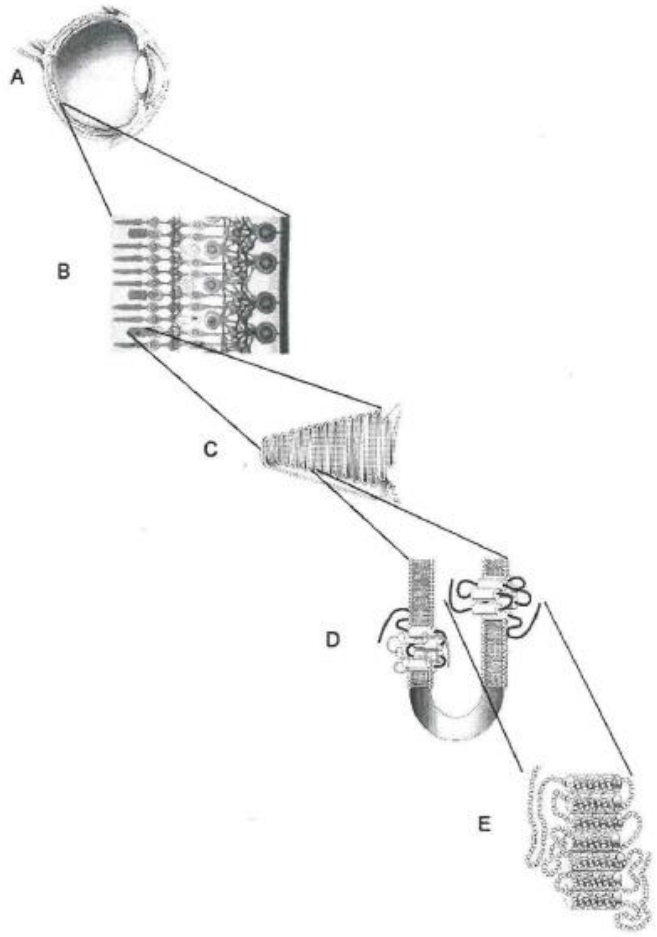

Figure 22.1. (A) The eye. (B) A Cross-section of the retina showing the principal cell types (including the rods and cones). (C) The outer segment of a cone. (D) Photopigments embedded in the membrane of the cone outer segment. (E) The amino acid chain of a cone photopigment.

Figure 2.1: The Eye

Looking at this picture (of the eye), one can hardly dispute the practical medical benefits of reductionism applied with care. Why can't environmental ethics too benefit from such information in careful, limited contexts?

[30] K. Aizawa and C. Gillett, 'Levels, Individual Variation, and Massive Multiple Realisation in Neurobiology', in *The Oxford Handbook of Philosophy and Neuroscience*, ed. John Bickle (New York: Oxford University Press, 2009), 544.

To show that it can, I explore various scientific and philosophical enterprises concerning the nature of the human brain. I look at evolution, moral evolution, evolutionary psychology, genetics, neuroscience and cultural evolution. This kind of multi-levelled approach is perfectly at home with the reductionist methodology adopted in this thesis. This methodology is not an *eliminative* reductionism, which 'seeks to eliminate one set of concepts and replace it with another one that is regarded as superior, perhaps in virtue of being more fundamental in some sense'.[31] Instead my approach is akin to that of Mary Midgley. Midgley is a naturalist and finds reductionism often a useful strategy (for instance, she develops an evolutionary theory for understanding emotions), however she rejects the eliminative reductionism involved in theories like Richard Dawkins, who would explain everything, and at all levels, in terms of 'selfish genes'.

Let us look to how methodological reductionism is applied in the biological sciences. Here there are two general reductive approaches. The first is reverse engineering. Reverse engineering is the dismantling of an object into its parts so as to determine from the features of those parts how the object works. This conception and approach is commonly deployed with man-made objects, but can be applied to biological objects. We can see this by looking at the eye in terms of its component parts (fig 2.1). This kind of reductive approach, although far more problematic when it comes to the brain, has obvious value. A second reductive approach used in biology is 'unrelated process theorising'; this refers to the utilisation of an unrelated process (neither downstream nor upstream) to explain an unknown phenomenon. For example, stating that the mind operates as a computer is an exercise in unrelated process reductionism—reducing the phenomenon's mechanics to that of the unrelated process.

Reverse engineering theories are fraught with possible pitfalls so it is worth considering the difficulties encountered by psychology when it comes to behavioural analysis. Points a. to c. below are summaries of such issues as collated through a lifetime of research in this area by neuroscientist William Uttal:

[31] Brian H. Baxter, 'Naturalism and Environmentalism: A Reply to Hinchman', *Environmental Values* 15 (2006): 62.

a) Most psychological constructs are extremely difficult to define. Indeed, almost all the definitions in our psychological lexicon are circular, invoking one mentalist term to denote the meaning of another.[32]
b) Many psychological constructs do not represent real psychobiological entities, but rather are mainly manifestations of our experimental methods and theories. Although it may be necessary to define mental processes operationally, there is no certainty that such operational definitions correlate exactly, what are from the point of view of other criteria, real psychobiological processes, mechanisms and events.[33]
c) Mental action and processes are generally inaccessible and can therefore only be inferred. These inferences may not reflect processes actually ongoing on in the brain but may instead depend on researchers' *ad hoc* theories or predilections. Many mental activities, processes, or mechanisms presented as being very specifically defined are, in fact, what MacCorquodale and Meehl (1948) designated as "hypothetical constructs". Because behaviour is presumed by at least a few of us to be neutral with regard to the actual underlying neural or cognitive mechanisms, we may be attempting to link these mythical, ad hoc, invented, inferred, or hypothetical constructs with specific brain regions that, at best, influence but do not uniquely instantiate or represent them.[34]

As Jesse Prinz argues, taking these points on board doesn't mean that, in other ways, psychology has nothing helpful to offer when it comes to our ethical enterprises (note he is not responding to Uttal intentionally here):

> Psychology can be conducive to normative ends in a number of ways. First, psychology can help us get clear on what things we value. Sometimes our values are not obvious to us. For example, people are not very aware of what leads to wellbeing… Second, psychology can also be used to identify inconsistencies in our values. People may have double standards... Third, psychology can help us see places where our reasoning is unduly influenced by factors we consider morally insignificant (geographical distance may lead us to underestimate the moral urgency of assisting the victims of catastrophes in foreign lands). Fourthly,

[32] William Uttal, *The New Phrenology* (The Penguin Press, 2001), 26.
[33] Ibid.
[34] Ibid.

> psychology can help us identify precise cross-cultural differences in values, and that can aid in setting the agenda for developing evaluative compromises.[35]

However, Prinz's well-made points do not overturn or dispute Uttal's basic point concerning the difficulties of behavioural reduction. Tim Shallice, a professor of neuropsychology, lists six different types of brain system organisations that could produce identical behavioural results:[36]

1. Modular systems. The brain is organised into a cluster of semi-independent, functionally specific modules that, when damaged, produce well-defined behavioural deficits.
2. Coupled systems. The brain is composed of individual modules that strongly interact, but each have their own specific functions.
3. Systems having a continuous processing space. The brain appears to be representing specific functions, which turn out to be simply different points on a continuum.
4. Systems of overlapping processing regions. The brain is made up of overlapping pairs of modules, with both modules in each partially representing the particular function under investigation.
5. Systems of semi-modules. The brain is made up of strongly interacting modular regions, whose respective functions are really only statistical averages of the many inputs that impinge on them.
6. Distributed and multilevel systems. The brain is organised such that its functions are widely distributed at different vertical and horizontal levels.

These possibilities all point to the difficulties surrounding a causal understanding of behaviours as products of brain activity: either behaviours have too many potential sources, or the brain itself is too unknown to make any kind of meaningful connections. I accept these problems are real, but hopefully will show that we should not give up because of these difficulties. After all, they are in a sense distinct from criticism levelled at reductionism *itself* as they really concern complexity. While there are many psychological

[35] Jesse Prinz, 'Reply to Dwyer and Tiberius', in Sinnott-Armstrong, *Moral Psychology,* 1:439.
[36] Tim Shallice, *From Neuropsychology to Mental Structure* (Cambridge, MA: Cambridge University Press, 1988), 249.

half-truths floating around that have arisen through the poor use of reductionism, there are likely just as many half-truths arising from holistic descriptions (drawing inferences from systems-level accounts).

The next point is that we should understand that reductionism applied to the challenge of uncovering the origin(s) of behaviours does not itself mean endorsing methodological individualism:

> Methodological individualism dominates economics, much of sociology, and all of psychology's excursions into organizational theory. This is the dogma that all human social group processes are to be explained by laws of individual behaviour, that groups and social organizations have no ontological reality, and that where used, references to organizations, for example, are but convenient summaries of individual behaviour.[37]

I agree we must pursue reductionism in conceptually the cleanest way possible, but we must also accept that cleanliness is next to godliness, and we are not in the fortunate position of the all-knowing gods. Reductionism allows us to understand the parts of wholes and their forms, as well as their influences, as properties of the whole**.**

Over the last decade scientific reductionism has taken on the lessons of philosophers and incorporated quite technical instructions for deducing causal relations from types of compositional relations (Table 2.1)[38]. Even so, we cannot avoid consideration of the central philosophical challenge on the table. This is the so-called Mereological Fallacy (MF), and it allows us to further refine the reductionism we need to for this project to work.

Table 2.1: Compositional Relations in the Sciences

Type of Entity as Relata	Compositional Relation
Processes	Lower-level processes implement a higher-level process
Individuals	Lower-level individuals constitute a higher-level individual
Properties	Lower-level properties realise a higher-level property

[37] D. T. Campbell, 'How Individual and Face-to-Face Group Selection Undermine Firm Selection in Organizational Evolution'. In *Evolutionary Dynamics of Organizations*, ed. J. A. C. Baum and J. V. Singh (New York: Oxford University Press, 1994).
[38] Aizawa and Gillett, 'Levels', 543.

Powers	Lower-level powers comprise a higher-level power

2.2.2 The Mereological Fallacy

Even some supporters of the MF have concluded that it is not technically a fallacy because it is not necessarily false to ascribe psychological predicates to parts of an animal that only apply to the behaving animal as a whole:

> If someone commits the mereological fallacy, then he ascribes psychological predicates to parts of an animal that apply only to the (behaving) animal as a whole. This incoherence is not strictly speaking a fallacy, i.e. an invalid argument, since it is not an argument but an illicit predication. However, it leads to invalid inferences and arguments, and so can loosely be called a fallacy.[39]

With this caveat made, we can see that MF is in the spirit of Wittgenstein's opposition to eliminative reductionism. As applied to neuroscience the MF states that we cannot ascribe properties of a whole human to its parts – centrally of course, to the brain. As applied to neuroscience MF was made famous following John Searle and Daniel Dennett's response to a book by M. R. Bennett and P. M. S. Hacker, *The Philosophical Foundations of Neuroscience,*[40] and the subsequent rebuttals from both sides.

Hacker and Bennett take the MF seriously, and reason that reductionist-based descriptions in this area of literature (consciousness, knowledge, neuroscientific research etc.) are false if they ascribe properties of the whole to a particular component of the whole. Thus Hacker and Bennett claim that one is not 'conscious with one's brain any more than one walks with one's brain'[41], and 'it is an illusion to suppose that anything whatsoever is added by ascribing knowledge, perception, and linguistic understanding (sort of, or otherwise) to the hemispheres of the brain'.[42] Their claim is that by doing this neuroscientists are, ironically, opting for a dualism of the Cartesian kind:

[39] H. Smit and P. Hacker, 'Seven Misconceptions about the Merelogical Fallacy: A Compilation for the Perplexed', *Erkenntnis* 79, no. 5 (2014): 1077–1109.
[40] M. R. Bennett and P. M. S. Hacker, *The Philosophical Foundations of Neuroscience* (Blackwell, 2003), xiv.
[41] M. R. Bennett and P. M. S. Hacker, 'Reply to Professor Dennett and Professor Searle', accessed 15 June 2015, http://info.sjc.ox.ac.uk/scr/hacker/docs/Reply%20to%20Dennett%20and%20Searle.pdf, 4.
[42] Ibid., 16.

> One major reason why we wrote our book was the firm belief that contemporary neuroscientists, and many philosophers too, still stand in the long, dark shadow of Descartes. For while rejecting the immaterial substance of the Cartesian mind, they transfer the attributes of the Cartesian mind to the human brain instead, leaving intact the whole misconceived structure of the Cartesian conception of the relationship between mind and body. What we were advocating was that neuroscientists, and even philosophers, leave the Cartesian shadow lands and seek out the Aristotelian sunlight, where one can see so much better.[43]

Hacker and Bennett's case certainly has implications for how neuroscientists present their findings, and one can easily agree that science is 'no more immune to conceptual error and confusion than any other form of intellectual endeavour'.[44] However, from the perspective of an ongoing research program Bennett and Hacker's criticism should be seen not as a restriction on empirical and reductive inquiry, but as a demand on (and for) a final and complete theoretical presentation of the research findings.

Obviously reductionism as applied to a whole is, in the first instance, intended to reveal that the whole is made up of parts. The MF says parts cannot be ascribed psychological predicates that only apply to the whole, but still it is obvious that parts of the whole differ (for instance, the brain is different from the liver, and different parts of the brain seem to have different functions, etc.), and these differences are of interest empirically. It is very useful, given the distinct properties of a part, to ascribe the properties of the whole to that part, if that part is of primary causal relevance to the operations of the whole. How, for instance, would a mechanic fix your car from always turning left, if you could not say turning left was a property of the steering drive? In such cases it would seem that in order to communicate clearly so that we might have our car drive as it should, we are well advised to be 'conceptually' incorrect. Thus our mechanic might diagnose the steering problem by saying 'it is the rack' that is 'pulling the pinion left'. So too, I suggest, with neuroscience. A point reinforced when we consider how much neuroscientific investigations have managed to achieve, and continue to achieve. Even Hacker and Bennett explicitly state that 'neuroscience contributes to explanations of irrational

[43] Ibid., 14.
[44] Bennett and Hacker, *Philosophical Foundations*, 4.

action',[45] and it seems a matter of explanatory symmetry that it may do so with rational actions too.

2.2.3 A Matter of Ontology?

I think that MF has real force against *eliminative* reductionism, but not all reductionism is eliminative. Methodological reductionism may be non-eliminativist, for it may allow for *emergent* properties. The emergence claim is that you may arrange parts and get extra properties of the whole, such as that of being a sphere, or the colour red. You arrange cogs and you get a turning mechanism. You arrange biological life and you get consciousness. Rearrange these parts and you don't get these emergent properties. So (for instance) consciousness is an arrangement of parts in a way such that if you removed a part it (consciousness) would cease to exist. The way of identifying those parts and how they function as to produce consciousness, which has certain properties – properties like reasoning – is via *non-eliminative reductionism*. This reductionism has no difficulty at all with the common putatively anti-reductionist claim that 'neuroscience is only a portion of the coupled brain-body-environment system'.[46]

As I see it, reductionism is a conceptual view of the ontological structure of existence, useful to appreciate the causality in any given phenomenon that arises across levels. A 'mechanistic explanation', with relationships of manipulability across these level, is a valid way of specifying causal relations between parts of interconnected wholes. Reductionism here allows us to understand how the interaction of parts causes change in the whole.

Some philosophers are tempted to make this point in terms of explanatory pluralism[47] and to think it means rejecting non-eliminative reductionism, but this is to confuse the metaphysics of causal reductionism with the realisation of it. We should be particularly careful when neuroscience makes empirical causal claims,[48] often in a language that, like our mechanics', is conceptually 'incorrect', owing to the difficulty of piercing through so many interacting levels to uncover a causal string for a given phenomenon. But this does

[45] Ibid., 365.
[46] C. Heyser and A. Chemero, 'Methodology and Reduction in the Behavioural Neurosciences: Object Exploration as a Case Study', in Bickle, *Oxford Handbook*, 86.
[47] Ibid., 87.
[48] John Bickle and A. J. Silva, 'The Science of Research and the Search for Molecular Mechanisms of Cognitive Functions', in Bickle, *Oxford Handbook*, 92.

not rule out the reality of this causation, nor that knowledge of it might not be very useful as it is with our car fixing mechanic.

In Section 2.7, I will explain what a *relationship of manipulability* is, and define acceptable criteria for deducing a theory of reduction among incomplete sciences ranging 'across levels'.

2.3 Universal Causal Determinism

It is likely with any kind of naturalism, and perhaps especially a reductionist (if non-eliminativist) naturalism, that questions of determinism and fatalism are going to be lurking in the background, for the arguments from naturalism and physical sciences that lead to universal causal determinism are strong. In fact, I favour universal causal determinism (UCD), as I do not think accepting this position is as controversial as many suppose. Perhaps, as I suggested earlier, we place too much emphasis on our 'free choice'. In other words, our ideas about how to solve problems and our capacity to actually solve them may be exaggerated and certainly overestimated; an exaggeration and overestimation that hinders us in the search for effective ways to achieve our various kinds of ends.

If UCD is correct then X-future is a necessary product of the prior state of the universe Y. This necessity would seem to imply that there is no freedom, therefore no human freedom, when it comes to manipulating the present for alternative futures; here determinism likely implies fatalism. What then of ethics, and what of strategic changes to further the effective implementation of such ends? Under UCD, nothing can change except for the universe changing, and the universe changes only as determined under the general laws. This doesn't mean (obviously) that we might not recommend 'change', but we do so only as part of the deterministic universe that produces 'humans making changes'. It is not 'change' in the sense that we can actually manipulate this unfolding universal construction and alter its direction through time. No-one controls unfolding reality, but we do have thoughts and move within it.

Now of course if UCD is true then it is no response to say, as did American psychologist and social philosopher, B.F. Skinner, that an elite can yet manipulate the causal impacts on

the non-elite masses; there is no room to introduce 'causally appreciative masters', precisely because they too are subject to the same causal necessity. If within a UCD reality a human is born with an enlarged prefrontal cortex (PFC) that facilitates improved causal analysis, and this has a survivalist advantage in a given environmental niche, then an 'elite' might emerge in accordance with standard evolutionary theory by natural selection. But this is not (and given the explanation in terms of natural selection, can't be) because those 'elite' individuals have any power or capacity to stand against natural necessity.

If we hold the view that the world is deterministic then I am writing as B. F. Skinner's critic J. E. R Staddon argued, 'without choice/freewill, due to such conditions being the nature of reality itself'. At this point, Kirkman makes the pertinent remark that:

> Something has to give here. In order to arrive at a coherent account of human morality, empiricists must either find a more sophisticated way to incorporate the possibility of rational deliberation and free choice, or they must purge moral language of any reference to will or self-command.[49]

But while clearly pertinent, the remark itself seems to me to involve a mistake. To propose that we must 'purge moral language' is over the top, and indeed would be foolhardy, precisely because such language appears to have serve a useful practical purpose, as does the capacity to reason itself. What we have – and we certainly do have moral language in a set of embedded 'language games' comprising our 'forms of life' – rests on and developed through our natural, evolutionary, history. What we require is, as Lewis Hichman argues, a 'Darwinian Humanism': 'Investigate the evolutionary sources of human behaviour without *reducing* the internal, symbolically mediated experiences of human affairs to an external, mechanistic series of explanations'.[50]

While philosophical concerns with UCD may seem pressing, it is usual for such concerns to be overlooked or ignored in a naturalist research programme, and this is what I will do here. Big picture problem of ultimate determinism should not prevent local causal

[49] Kirkman, 'Darwinian Humanism', 11.
[50] Lewis Hinchman, 'Is Environmentalism a Humanism?' *Environmental Values* 13 (2004): 20.

discoveries from aiding us in our decision-making; whether this flows from actual human-directed change or is just another part of universal change.

2.4 Naturalism and Environmental Ethics

> *If the history of the twentieth century has anything to teach, it is that secular ideologies are even worse than religious creeds at fomenting cruelty and violence. This leaves only one possible source for a new moral code-natural law, the law 'written on the tablets of eternity'.*[51]

Naturalism has a *metaphysical* interpretation in which all facts are ultimately natural, including moral facts, as well as an *epistemological* interpretation (which some refer to as *methodological naturalism*) that holds all possible forms of enquiry are continuous with the empirical worldview. Patricia Churchland's view of (a proper) naturalism is good starting point: 'a monist ontology and a non-mystical, realistic, materialistic and naturalistic philosophical outlook'.[52] On this view the mind is a function, process or manifestation of the information-processing mechanisms of the biological brain:

> Naturalism, while shunning stupid inferences, does nevertheless find the roots of morality in how we are, what we care about, and what matters to us—in our nature. Neither supernaturalism (the other worldly gods), nor some rarefied, unrealistic concept of reason, explains the moral motherboard.[53]

As Lynch points out, naturalism is hard to define in terms of what it is that positively counts as natural prior to what we allow and use in specifying this or that 'natural' theory, however it can be easily defined negatively, in terms of what it excludes, which is, of course, the super or supra-natural.[54] This is actually the view most scientists take as defining their naturalism: a grave fear of any mystical or god-like entity interfering or intruding into the world we seek to understand in reflectively secure terms. So this is a reasonable starting point on which to build a more robust form of naturalism.

---

[51] Ophuls, *Plato's Revenge,* 20.
[52] Uttal, *New Phrenology*, 4.
[53] Churchland, *Braintrust*, 6.
[54] Tony Lynch, 'Winnicott's World: The Mystical Foundations of Reality', *Transcendent Philosophy* 4, no. 4 (2003).

Such a robust naturalism – indeed, one I think too robust in its easy collapse of 'ought' into 'is' – is to be found in the practical environmentalism of B.G. Norton**:**

> Attempts to separate factual from value content in the process of deliberation are rejected; there is only one method for evaluating human assertions, including assertions with all kinds of mixes of descriptive and prescriptive content, and that is the method of experience—active experimentation, when possible, and careful observation otherwise. The scientific method is embraced as the best approach to evaluating hypotheses about cause and effect, but also about what is valuable to individuals and cultures.[55]

Similarly, Willliam Ophuls, another environmentalist, or more precisely 'political ecologist', argues for an ethical basis grounded in lessons from nature itself: 'ecology, physics, and psychology, that is, biological nature, physical nature and human nature—reveal fundamental and eternally valid moral principles with which to reconstitute our polity'.[56]

Such approaches, as I suggested, too easily collapse the is-ought distinction (though any naturalism here will do so on the ultimate level, for if, as Kant said, *ought* implies *can*, then oughts are things for us humans, and we humans are biological historical constructions with certain, limited, capacities and with real incapacities), but they open our eyes to point of vital significance; for we cannot now see ourselves as independent of nature, subject to special creation, nor as in some metaphysical sense, dominant over nature. As the philosopher Zeno suggested we must live in agreement with nature. But this ultimate, constitutive, agreement is one thing, it is quite another thing to live sustainably in nature as a natural being, and this is something we know in so far as we know – our Anthropocene banality – that this is *not* what we are presently doing. As Ophuls insists:

> Human demand must… be moderated to a level that can be supported by the available flows, the capital stock that produces these flows must be persevered at all costs, and the remaining reservoirs of fossil capital should be conserved as

[55] Norton, 'Facts, Values, and Analogies', 65.
[56] Ophuls, *Plato's Revenge,* 22.

> much as possible. In other words, we cannot continue to behave as if we were the omnipotent lords and possessors of the universe.[57]

But how do we get there? Here I disagree with Ophuls, and diverge on how to use the naturalist account.

Ophuls argues that our civilisation has exalted the values of the rational-logical left brain over those values of artistic-intuitive right brain. He claims, as Pascal did, that we have overemphasised reason, the faculty associated with our left brain, while ignoring 'reasoning's of the heart', which he claims 'have their neurological seat in the right side of the brain'. Ophuls then (like Schiller before) claims that the remedy for this defect lies in in an aesthetic education.

This left/right schematic approach, as we will see in later chapters, gets it mostly wrong about what our best naturalist account of our brains tells us, and it has no practical merit other than to remind us of the place and power of our ethical drives, capacities and concerns.

Further, the left/right brain meme, with its 'rational'/'intuitive' divide, is itself a potentially misleading as it encourages us to place instrumental rationality on one side and (some form of) 'intrinsic valuation' on the other, and to do so in a competitive or exclusivist setting. In the context of environmental ethics Baird Callicott (2002) and Mark Sagoff (2004) have argued that environmentalists should play down instrumental arguments 'focusing on the most or cost-effective means to achieve a specific end', claiming this view ignores the *value* of that end. Bryan Norton's response is to ask how on earth practicality can be obtained unless one adopts an instrumental view:

> What concerns me about this trend is that it isolates environmental valuation and goal-setting from scientific discourse, creating ideological positions about the nature of value, polarized positions that cannot be resolved by experience or deliberation. To place this argument in historical context, Sagoff and Callicott seem to be reviving arguments against 'naturalism' in the study of environmental

[57] Ibid., 68.

> values and goals, relying on mysterious 'non-natural' qualities—qualities not visible or measurable by the senses.[58]

Norton's approach is self-described as a form of methodological naturalism (Section 2.2), and he argues:

> It does not contradict Hume (if understood as prohibiting the deduction of values from facts), nor need violate Moore's cautions against defining values as facts (in terms of natural qualities)… the changes need to support a new conservation consciousness are usually reorganisations and reconceptualizations of facts, not deductions from value neutral *facts*.[59]

We will discuss the mentioned fallacies in more detail shortly. We should also recognise that the naturalist approach runs counter to three classically held beliefs about ethics:

> First biological perspectives are inherently evil, as they create predetermined outcomes, thereby eliminating free will. As philosophers and psychologists, such as Daniel Dennett, Steven Pinker, and Daniel Wegner, have argued, nothing about an evolutionary or biological perspective leads inextricably to the notion of a determined, fixed, or immutable set of judgements. Biology doesn't work this way. Our biology, and the biology of all species on earth, sets up a range of possible behaviours. The range we observe is only a limited sampling of the potential variation. This is because our biology interacts with the environment and the environment is constantly changing... Second, if a biological perspective on morality is true the moral principles must be encoded in the DNA. Different amino acid sequences link to different deontological rules, some for harming and some for helping… Nothing in our genome codes for whether infanticide, incest, euthanasia, or cooperation are permissible and if permissible, with which individuals…Third, even if biology contributes something to our moral psychology, only religious faith and legal guidelines can prevent moral decay.[60]

[58] Norton, *Sustainability*, 66.
[59] Norton, 'Facts, Values, and Analogies', 67.
[60] M. D. Hauser, *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong* (Harper Collins, 2006), 456–457.

Doubtless ethics has a large cultural dimension just as does anything that, if potentiality is to be actualised, demands education, but this does not at all mean that culture floats free of its biological foundations (Wittgenstein himself pointed out that you can't get a cat to appreciate that you are *pointing* at this or that); nor does it mean naturalised accounts cannot provide *useful* information (see Section 2.7).

Further issues with the viability and adequacy of a naturalist account are to be found in Kirkman's strategic problems for environmental ethicists relying on an empiricist account:[61]

1. There is no perspective from which to criticise moral sentiments themselves, so the theory ends up validating whatever inclinations people already happen to have.
2. Darwinian accounts of moral sentiment in terms of heritable social instincts must allow for natural variation among individuals within the population. There seem to be no grounds on which to condemn such inclinations or the people who happen to have them except by appeal to public opinion.

In an investigation such as this we must be mindful of Point 1; it is very easy (as scientists such as Richard Dawkins and Stephen Hawking have shown) to over-extend scientific theory beyond its objective scope by utilising it to bolster personal opinions on metaphysical issues. Point 2 is related to Point 1, for the 'objectivity' of such personal inclinations might seem possible only on the basis of 'public opinion'. However, I see no reason why we cannot accept both these strategic problems as warnings against personal bias, partiality and prejudice, and against the dangers of herd thinking, rather than anything deeper.

Those environmental ethicists who *do* incorporate evolutionary theory are likely to do so in a search for a 'metaparadigm'. For example, Donato Bergandi attempts to identify some shared ontological and ethical foundations that he believes make it possible to 'distinguish a minimal common basis for environmental ethics':

[61] Kirkman, 'Darwinian Humanism', 8.

> Humans, like other species, are members of the earth's single biotic community; Humans like other species, are an integral part of a system of evolutionary and ecological, biotic and abiotic relationships that allows them to survive, proliferate, and develop to the best of their potentialities; Humans must control their proliferation and their economic development in a way that allows for the highest possible level of biodiversity and evolutionary potential on the planet.[62]

This meta-paradigm approach – with its too easy slide from is to ought in the third sub-sentence – seems congenial to Norton and Olphus, but it is not my approach here. I am not trying to argue that 'ecological interconnectedness' gives rise to moral obligations to the environment, rather I am looking at our Anthropocene awareness of the unsustainability of our present practices and asking why we have not managed to give proper practical realisation to this realisation in terms of an effective ethic of sustainability? And I am asking this in terms of a causal understanding our thinking and decision making, and ultimately our behaviour.

2.4.1 Alternatives to Naturalism

> *The tension arises between two distinct strands in the history of modern philosophical ethics: an empiricist strand that runs from Hobbes to Hume, through Darwin and on down to present-day socio-biology and evolutionary ethics, and a humanist strand that runs from Rousseau, through Kant and on down to Sartre and other defenders of human dignity.*[63]

Thus far we have looked at some of the challenges to the use of the naturalist and reductionist methodologies, but we have not considered the proposed alternatives. I will briefly explore two alternatives with similar roots; *Humanism* and *Expressivism*.

Humanism has been defined as 'any view that emphasises the autonomy and dignity of humans as free moral agents and our capacity for self-improvement'.[64] As such it is often taken to imply a rejection of naturalism, biologicalism, and behaviouristic social sciences.[65] Kant famously promoted humanism with his notion of that transcendent

[62] Bergandi, *Structural Links*, 21.
[63] Kirkman, 'Darwinian Humanism', 4.
[64] Ibid., 5.
[65] Hinchman, 'Is Environmentalism a Humanism?', 12.

rational freedom we are supposed to possess as centres of 'pure practical reason'. This conception of our ultimate and distinctive freedom is at the core of humanism:

> Human freedom is the freedom to make up stories about ourselves, constrained only by our own imaginations and what it takes to get others to listen and accept them as a form of 'truth'. Humanism thus seems to aim to preserve a symbolically-rich, freedom-conferring, endlessly-changing, world of human narrative, embodying meanings and values.[66]

As a non-eliminativist reductionist I have no bone to pick with humanism as such, but when it comes to issues of the applicability and implementation of ethical strategies the vital challenge is to provide an objective account of how the brain works when it comes to such deliberations; and it is here that humanism's propensity to human exceptionalism can get in the way. More than this, an (over)commitment to humanist 'narrative freedom' may blind us to the adverse real consequences of what we may see as a plain humanistic good (e.g. pushing for renewable energies may end up causing more environmental damage than a bottlenecked oil supply). Here the trouble is an obsessive and eclusivist focus on humanist narratives. As Brian Baxter writes of Hinchman, who he takes to be just such a humanist:

> What Hinchman appears to be against is not reification, objectification, reductionism and determinism, but the replacement of hermeneutics with any other mode of explanation of human behaviour whatsoever, whether that be reifying or non-reifying, reductionist or non-reductionist, determinist or non-determinist. What he seems to support is a picture of human life as largely dependent upon stories, which people tell about themselves, which have unique meaning for them, and in terms of which they find their meaning and identity. He is happy to let this story-telling approach be spread beyond human life to encompass the non-human. What he seems unwilling to countenance is anything that replaces, or even supplements the stories.[67]

[66] Baxter, 'Naturalism and Environmentalism', 57.
[67] Ibid., 57.

Expressivism, itself a humanist variation, arguably offers more than naturalism when it comes to understanding moral discourse:

> Gibbard and Blackburn argue that expressivism affords a better explanation of crucial features of moral discourse and practice than ethical naturalism, namely, (1) the non-hypothetical character of moral judgment, (2) the link between moral judgment and motivation, (3) the nonrelativistic character of moral assessment, and (4) the resistance of fundamental moral disagreement to resolution by empirical means. They account for (1) – (3) by providing a semantic theory according to which moral judgments express non-hypothetical, motivating attitudes on the part of the speaker.[68]

If my concern was with the semantics of moral expression then expressivism would be of great interest. But this is not my concern. My concern is with the practicality and effective implementation strategies needed to meet the Anthropocene sustainability challenge. This concern means seeing how far a naturalistic understanding of our ethical lives (capacities, incapacities, potentialities, limits, and so forth) can take us. I am not saying, 'use naturalism, reductionism and determinism to arrive at an understanding of meaning in one's life'. I am simply saying that when it comes to environmental ethics, in the Anthropocene, let us see how far this line of thought can take us in complementing traditional humanistic and linguistic accounts of environmental ethics.

In short, I have selected a methodology I think works best for a particular case: responding to environmental crisis in terms of informing potential frameworks for action. We will be better off here, as Norton suggests, if we work hand in hand with the sciences. But as we will find in 2.6 we must be careful not to be overly pragmatic.

2.4.2 Naturalistic Fallacy and Hume's *Is-Ought* Distinction

> *... for the good of the species, the government ought not to interfere with nature's tendency to let the strong dominate the weak.*[69]

[68] P. Railton, 'Naturalism Relativized', in Sinnott-Armstrong, *Moral Psychology*, 1:38.
[69] Herbert Spencer, *Data of Ethics* (1883; repr., Belle Fourche, SD: Kessinger, 1998).

> *... the negligence or perverseness of mankind cannot be excused, if their discourses on morality be not much more clear than those in natural philosophy.*[70]

Taking aim at Herbert Spencer, what is now referred to as the 'naturalistic fallacy', was named by G.E Moore in 1903 in his *Principia Ethica*, and established by the 'Open Question Argument'. For our purposes, Churchland's summary statement is sufficient:

> There is no answer to the question of what natural properties are identical with the good or the right or the valuable. That is because for any proposal, the open question [but is x good/bad?] can always be asked.[71]

While Richard Joyce has suggested that the open question argument is 'based on the confused views of necessity, a prioricity, and analyticity which dogged early-twentieth-century philosophy',[72] there is sense to Moore's point: Any property – and it does not have to be one a physicalist would approve, just any property at all – cannot be identical to/capture the meaning of 'good', for of any such property it makes sense (semantically) to say 'But is property X good?' In the context of our concerns, the point is that while science might tell us why we are prone to a certain kind of evaluation (i.e. ranking our children ahead of others') by pointing to the biological and evolutionary centrality of genetic lineage, it cannot tell us that such a concern is a genuine ethical value:

> Whereas neuroscience might be able to identify the neurophysical correlates for evaluate notions such as preferences and attitudes, lying and the distinction between in control and out of control behaviour, neuroscience cannot, in and by itself, provide the basis for their evaluation.[73]

Of course one might deny the former claim anyway. How could such things as 'lying' be specified by drawing on neurological facts that are true of an isolated individual?

---

[70] John Locke, *Essay Concerning Human Understanding* (1689; repr., New York: Dover, 1959), III: xi, 17.
[71] Churchland, *Braintrust*, 187.
[72] Richard Joyce, *The Evolution of Morality* (Cambridge: Massachusetts Institute of Technology Press, 2006), 152.
[73] T. Buller, 'What Can Neuroscience Contribute to Ethics?' *Journal of Medical Ethics* 32 (2006): 63.

I have already discussed a specific way in which facts are relevant to oughts (when we consider 'can and cannot'), Churchland, however, questions the validity of the is-ought distinction, largely because on a more general level she sees the role of naturalism as more significant for applied ethics than philosophers have previously credited:

> Naturalism in ethics should no longer be hobbled by the dictum that you cannot infer an ought from an is. What you can do, however, is come to a decision about what you ought to do without relying on any normative rules or maxims. That is what humans, and undoubtedly other animals, in fact do. From this perspective, many new questions in ethics arise. [74]

My point is not to endorse Churchland on this, but with her to insist that naturalism does have a greater role to play than often thought when it comes to thinking about the implementation of ethical value strategies.

My concern is with the implantation strategies for Anthropocene sustainability, in particular, with the executable step of Traer's 'Doing' process. As I have said, Kant said 'ought implies can'. This means that any genuine *ought*, as opposed to something in the optative mood, must be informed by, or connect with and into an *is* centred on our human nature, and especially on that nature as embodied in (the structures and processes) of our brain. It may be that 'what my parents taught me' is crucial to understanding my ethical commitments, but even beneath that, and on both the sides of the pedagogue and the neophyte, there the question and facts of evolved neurospychological arrangements. It is here, ultimately, that 'is' comes into contact with 'can and cannot'.

I will not be replacing normative questions with scientific ones, nor violating the is-ought distinction. My task is to deploy a non-eliminative reductionist naturalism so as to determine what options we might have for achieving the sustainable realization of Anthropocene sustainability (the simple banality).

To help explain, consider briefly the success of this practical methodology in Daniel Levitin's understanding of the effects of music on the brain, and ask yourself why should

[74] Patricia S. Churchland, 'Inference to the Best Decision', in Bickle, *Oxford Handbook*, 427.

not illuminate environmental ethics in a similar fashion if we have the factual information to do so? Why not use facts about our socio-biology to aid in the application of our environmental ethics, just as it does when it comes to our use(s) of music?

> The power of music to evoke emotions is harnessed by advertising executives, filmmakers, military commanders, and mothers. Advertisers use music to make a soft drink, beer, running shoe, or car seem more hip than their competitors. Film directors use music to tell us how to feel about scenes that otherwise might be ambiguous, or to augment our feelings at particularly dramatic moments. Think of a typical chase scene in an action film, or the music that might accompany a lone woman climbing a staircase in a dark old mansion: Music is being used to manipulate our emotions, and we tend to accept, if not outright enjoy, the power of music to make us experience these different feelings.[75]

2.4.3 A Sufficient Naturalism: Ontology, Epistemology and Environmental Ethics

Although I have spoken in broad strokes about naturalism, its use and interpretation, what is also required is to present a form of it which can be used to effectively operationalise an environmental ethic for the Anthropocene. To do this we need a scientifically uncontentious and straightforward ontology and epistemology. Let us start with an ontological position and apply it to naturalism, as Flanagan, Sarkissian, and Wong have in defining *ontological naturalism* as:

> the view that what there is, and all there is, is natural. Everything that exists/has existed, happens/will ever happen, is a natural phenomenon, process, or event. Every property, event process and thing, if it *genuinely* exists/is happening, did exist/happen, or will exist/happen, is natural.[76]

In ontology something has an objective reality if its existence does not depend on it being experienced. A key indicator of ontologically objective phenomena is the possibility of independently accurate measurements of its properties. In contrast, something has a subjective reality if its existence consists in it being experienced. (When I stand in front of

[75] Daniel J. Levitin, *This is Your Brain on Music* (New York, NY: Penguin, 2007), 9.
[76] Flanagan, Sarkissian, and Wong, 'Naturalizing Ethics', 21.

my ontologically objective tree in my backyard, I have a tree-ish experience.) In the case of the former, the appearance/reality distinction has a place, in the case of the latter it does not, for an illusion or delusion is still a fact of phenomenological awareness/consciousness**.**

Crucially, this simple ontological distinction does not imply substance dualism. Like many others, I hold that the ontologically subjective is created 'in' or by the ontologically objective brain: 'that all mental activity occurs as a result of brain activity cannot be denied',[77] and 'consciousness is entirely caused by neurobiological processes and is realized in brain structures'.[78] Understanding this causality, we should still recognise that both the ontological objective (the natural world), and the ontologically subjective (our experiences of it), are real, even if logically one involves appearance/reality distinctions, and the other does not.

A specific neural activity occurs in the brain when you see the colour green. That activity is not green itself. This might lead someone to say that the 'looking green' is an illusion somehow produced by brain activity. 'Illusion', however, has the wrong connotation; the mind is a direct causal product of that neuronal action which occurs as a result of interaction with the physical universe. There is no dualism here. We just do not know precisely how this 'green experiencing' is created by a brain in a causal world, though we assume that it must be so created. If we understand what environmental conditions stimulate the brain we can understand the causal origin of the experiences without understanding exactly how the brain manifests these causal impacts in subjective experience(s).[79]

In epistemology, we make truth claims. A claim has epistemological objective truth if:

1. Ontologically objective data supports the claim. For example, the claim that cancer kills people is supported by observed medical evidence independently verified over and again. This information I consider the highest-ranking support for a truth claim.

[77] Uttal, *New Phrenology,* 35.
[78] John R. Searle, 'Consciousness', in *A Cognitive Science Dialogue: Consciousness East and West* (2005), 1.
[79] The latest suggestion is through an area referred to as the *Clustrum* as one of the key 'conductors of consciousness' (F. C. Crick and C. Koch, 'What is the Function of the Claustrum?' *Philosophical Transactions of the Royal Society, Biological Sciences* 360 (2005): 1271–1279.)

2. If many ontologically subjective states concur. For example, 'we all agreed that the earth must be square because we experienced the land as flat'. This information has the second highest ranking until such time as ontologically objective data in (1) overturns it.
3. Truth is 'subjective' if all it draws on is individual (non-shared) experience(s) (My friend swears he saw a ghost last night). This has the lowest epistemological ranking, until such time as (2) and ideally (1) can be demonstrated: we all saw the ghost the next evening, we then captured it and each took the same measurements of it.

2.5 Evolution and Environmental Ethics

> *Evolution might favour people who are nice to their neighbours, but it might also favour people with genocidal tendencies, for the same underlying reason. Thus if you're looking to evolution for moral truth, you're barking up the wrong tree.*[80]

> *Science can advance ethics by revealing the hidden inner workings of our moral judgments, especially the ones we make intuitively. Once those inner workings are revealed we may have less confidence in some of our judgments and the ethical theories that are (explicitly or implicitly) based on them.*[81]

If the crux of naturalism is biology, the crux of biology is evolutionary theory. We need, then, to think about how, when it comes to ethics, one uses evolutionary theory.

Some want to derive from evolution an objective morality. This has seemed an alluring prize for many neuroscientists, biologists and philosophers. Such attempts have drawn on evolutionary theory in a variety of ways, but none yet have been successful, and pretty much all such efforts have been punished in the philosophical realm. Sam Harris offers a particular array of faults in *The Moral Landscape*, and Joshua Greene in *Moral Tribes*, as the above quotations suggest, dances along lines perhaps finer than he realised at the time.

[80] Greene, *Moral Tribes*, 186.
[81] Joshua Greene, 'Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics', *Ethics* 124, no. 4 (2014): 696.

We must then be very clear when we think about how evolutionary theory and ethics do, and do not, connect. Here William Fitzpatrick is useful.

1. Descriptive Evolutionary Ethics: appeals to evolutionary theory in the scientific explanation of the origins of certain human capacities, tendencies, or patterns of thought, feeling and behaviour. (e.g.: Relying on a module of evolutionary psychology to explain behaviour.)
1. Prescriptive Evolutionary Ethics: appeals to evolutionary theory in justifying or undermining certain normative ethical claims or theories. For example: the appeal to evolutionary theory to justify free market capitalism or male-dominant social structures.
2. Evolutionary Metaethics: appeals to evolutionary theory in supporting or undermining various metaethical theories. For example … to undermine the claim that there are objective moral values, or to cast doubt on whether we could have justified beliefs about such values.[82]

With regards to 1), I intend to reveal relevant descriptive facts, of a sociobiological kind, that may aid in understanding environmental ethics. With regards to 2), in no way whatsoever will I appeal to evolutionary origins, or scientific facts regarding the operational context of the human brain, to *justify* outlandish moral views. My approach rests, as I have said, on the biologically unsurprising value to us of continuing and sustainable human civilizational life.[83]

I am not, therefore, taking a prescriptive approach (at least not in this evolutionary capacity). Bernard Williams argues that the role of socio-biology when it comes to ethics is not so much a matter of drawing out or justifying positive prescriptions, but is rather negative and limiting.[84] For instance, given what we know about human nature there may be certain proposals we know are useless or bound to fail horribly (e.g. Plato's idea that in

[82] William Fitzpatrick, 'Morality and Evolutionary Biology', in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford University, Summer 2012), http://plato.stanford.edu/archives/sum2012/entries/morality-biology.

[83] One could take the banality further and argue that a concern for a sustainable environment (one in which one can live and reproduce) is an evolutionary given, and one that centrally and intrinsically connects with human welfare. From this position one could contend that this environment/welfare is constitutive of the possibility of human flourishing. And one could argue that any possible (not self-eliminating) moral system must rest on and respect this tripod (environment, welfare, flourishing).

[84] Bernard Williams, *Ethics and the Limits of Philosophy* (London: Routledge Classics, 2006), 146–172.

the ideal ethical state every child would count every adult as 'Mum' or 'Dad', and every 'Mum' and 'Dad" every child as 'their own'). Here sociobiological facts plausibly constrain ethical possibilities, as Aristotle himself pointed out in his response to Plato.[85]

Yet even with this example, some may still be uncomfortable. For instance, they may question whether an ethic really is 'biologically impossible', pointing to the extraordinary variety of anthropological arrangements we see when it comes to kinship, though none, it has to be said, look much like Plato's proposal.

With regards to 3) I will be keeping an open mind, though obviously I am deeply sympathetic to the naturalist position that 'moral philosophy should not employ a distinctive *a priori* method of yielding substantive, self-evident and foundational truths from pure conceptual analysis'.[86]

The point of drawing on science here is not to challenge our normative views themselves, but rather to inform us of their natural origin(s) and their place and role in the economy of human life.

Finally, we should also consider reverse lines of reasoning. For example, the argument that since it seems that evolutionary theory undermines our moral ethics, but our moral ethics can't be so meaningless, evolutionary theory must itself be flawed. Although this is an interesting conclusion, dismissing evolutionary theory on this basis isn't empirically justified and further it relies on an interpretation/conception of 'meaninglessness' that is quite subjective. A similar argument states that either the ways of reasoning that have evolved for one purpose can validly be generalised to others, or they cannot. If the former holds then evolutionary considerations do not undermine ethics, if the latter then the theory of evolution is itself undermined so cannot itself undermine ethics. The survival value of some ethical considerations, but not those of environmental ethics, has been discussed and if we may validly generalize ways of reasoning, then we can reason about environmental ethics (not that this facilitates in itself motivation to this end). And as we will see later that Alvin Plantinga's argument, that 'if we cannot generalise reasoning validly then we cannot validly infer the theory of evolution', faces serious in empirically orientated studies like

[85] Ibid.
[86] Flanagan, Sarkissian, and Wong, 'Naturalizing Ethics', 5.

Peter Ulric Tse's, who provides an account of evolved 'cross-module binding', which would empirically account for generalised reasoning (so we can genralise).

2.6 Pragmatism and Environmental Ethics

> *Associated with this pluralization of our experience in dealing with others is a second aspect, a serene acceptance of the plurality of worldviews and perspectives. Once one enters the pluralistic world of the Pragmatist, and recognizes that reduction of all points of view to a single, authoritative and certainly correct view of reality will never happen, one adopts a new, less argumentative attitude. This attitude accepts the complexity of the world, and the many conceptual and theoretical tools humans have created to understand that complex world, and seeks cross-perspectival understanding, believing that it is reasonable to expect that one will learn something from alternative points of view, even ones one finds initially foreign or even abhorrent. As [Richard] Bernstein says: 'Here one begins with the assumption that the other has something to say to us and to contribute to our understanding' (1997, 399).*[87]

Norton's appreciation of pluralism within the context of pragmatism is, of course, one of the roots of pragmatism, others include anti-foundationalism, fallibilism, the social meaning of the self, and contingency.[88] Loosely put, *environmental* pragmatism is when one 'develops strategies by which environmental ethics can contribute to the resolution of practical environmental problems'.[89] This thesis is a work of environmental pragmatism and realism.

To understand this pragmatism let me consider the central challenge to pragmatism in environmental ethics. This challenge comes from the philosophical 'purists' who, like Callicott, see the pressing challenge as that of developing an entirely new or novel non-anthropocentric ethic of intrinsic value. In my view, as with Ben Minteer in his recent work, *Refounding Environmental Ethics,* this purism has had the unfortunate effect of helping to marginalise that ethics:

[87] Bryan G. Norton, 'Leopold, Hadley, and Darwin: Darwinian Epistemology, Truth, and Right', *Contemporary Pragmatism* 10, no. 1 (2013): 6.
[88] Ibid., 4.
[89] C. Palmer, 'An Overview of Environmental Ethics', in Pojmon and Pojmon, *Environmental Ethics,* 30.

> It would be difficult to mount a convincing argument that environmental ethics discourse has made a significant contribution to tackling… societal challenges or that it has played an important role in the environmental policy process or conservation planning and practice more generally.[90]

Minteer (like Norton) is trying to encourage a pragmatic movement in environmental ethics itself, a 'refounding'. My project is a little different. I am seeking to inform environmental ethics with specific kinds of facts relevant to seeing it successfully applied. To do this I embrace and extend causal facts rooted in our biology**.** To do this we need not challenge principled, 'purist' environmental ethics as Minteer does. My approach can even be placed in the service of 'purists' like Callicott, for we must not lose sight of our pre-theoretical intuitions. After all, this is where we start from. And certainly, as I have pointed out earlier, we need not be instrumental about our intrinsic values.

Minteer frames environmental pragmatism in terms of influencing policy, setting it against the principled environmental ethics of the intrinsic value kind; and over this has clashed with Callicott**.** Such open adversity tends to push each side into extremes. On one hand we must be sensitive to the demands of policy pragmatism, but on the other hand, and if that policy is to be effective, we must acknowledge our intrinsic evaluations, both of the value of our human existence and of that nature in which it has developed and has its place. In my view there is an underlying commonality to both these views when one thinks (as both do) of getting results. One needs a metric of assessment to determine if policy works, and to determine what working amounts to, one must have a set of values that respect Kant's 'ought implies can' condition. In general, Minteer's and Norton's naturalist pragmatism is missing this fundamental component: specifically, an understanding of the causality of human decision making from the sociobiological/evolutionary lens (including neuroscience). As Catherine Larrere points out the purist position rest on 'the general idea… that if we change our relationship to nature, by taking into consideration its moral dimension, we will behave both rightly and usefully'.[91] Equally obvious is the point that such behavioural changes must be available to naturally evolved beings we are.

[90] Ben A. Minteer, *Refounding Environmental Ethics: Pragmatism, Principle and Practice* (Philadelphia: Temple University Press, 2012), 2.
[91] Catherine Larrere, 'Two Philosophies of the Environmental Crisis', in Bergandi, *Structural Links*, 142.

2.7 Incomplete and Multileveled Sciences

A central problem with the science I explore in chapters 3 through 9 is that in many cases it is drawn from emerging fields that, as such, are as yet incomplete and largely unintegrated. On the other hand, it is precisely this breadth of approach which (I hope) enables the thesis to think across the biological and cultural spectrum of our interactions with particular environments as it gives rise to our behaviours as they impact on the natural environment**.**

This style of approach to explaining our morality is gaining prominence, and if it is to be fruitful it will (and must) allow for causal threads between levels of explanation:

> The co-development of genes and culture could be part of a multi-level explanation of morality… several levels of explanation may coexist in evolution, and human nature may be seen as a layered structure of genetic and cultural levels.[92]

Given that several levels of explanation may coexist in evolution, the following is a brief list I've collated of the ways in which evolutionary theory has been used:

a) To explain the biological changes over time in the reproducing organism, as Darwin first theorised.[93]
b) Given (a) to predict or 'just so' a likely evolutionary history of particular animals (including humans) given those facts we have about the past (thanks to the age and characteristics of fossils/rocks etc.) in light of present forms (what functionality animals have today). This may include a breakdown of the construction by particular environments of the brain's functionality (evolutionary psychology).[94]
c) To see evolution as a kind of total fact which, as such, implies an immanent and 'real' ethical imperative: 'Survival of the fittest'.[95]

[92] H. Looren de Jong, 'Evolutionary Psychology and Morality. Review Essay', *Ethical Theory and Moral Practice* 14, no. 5 (2011): 124.
[93] Charles Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (1850. Repr., New York: Signet Classics, 2003).
[94] John Tooby and Leda Cosmides, 'Conceptual Foundations of Evolutionary Psychology', in *The Handbook of Evolutionary Psychology*, ed. D. Buss, 5-67. Hoboken, NJ: Wiley, 2005.
[95] Spencer, *Data of Ethics*.

d) To argue that the fundamental unit, on which evolution ultimately operates is located at the interaction of genes or elsewhere[96] (which objects fulfil the proposed requirement of replication).[97]
e) To apply evolution as a mechanical theory to the minds ideas, essentially the idea that ideas evolve (i.e. Dawkins' memes theory).[98]
f) To apply the mechanics of evolution to culture (including ideas) to understand its history as in (b), i.e. cultural evolution and the potential future of cultures.[99] Note: (e) is distinct from (d) in that it does not require stringent alignment to biological evolution (existence of replicators etc.)
g) Neural Darwinism. Also referred to Evolutionary Epistemology, or Universal Darwinism. Key neural theory posited by Nobel Laureate Gerald Edelman. This theory seeks to explain ongoing developmental selection, and experiential selection via/within the brain. Experiential selection referring to synaptic selection between repertories of neural groups, analogue to species/organism level of selection.[100] Or as Universal Darwinist Henry Plotkin states: 'What is meant by Universal Darwinism is that both forms of evolution, between and within organisms conform to identical processes'.[101]

Given this profusion of projects and orientations, it is better to admit the provisionality of any research in this area, including my own. Yet history shows us such sciences can come together to give us powerful causal theories. So let us return to our initial discussion on reductionism, only this time we will take more seriously those in neuroscience who, at least in the last few years, have used it more carefully than a sceptical philosopher might appreciate.

Within neuroscience today there exists two forms of reductionism. 'Ruthless reductionists' seek to explain how mental processes are performed in the brain in purely molecular and cellular terms. 'Mechanistic reductionists', on the other hand, look to system-level explanations of brain composition, which involves identifying the roles of the operating

[96] Richard Dawkins, *The Selfish Gene* (New York, NY: Oxford University Press, 1976).
[97] David L. Hull, 'Units of Evolution: A Metaphysical Essay', in *The Philosophy of Evolution*, edited by U. J. Jensen and R. Harré, 23–44 (Brighton, England: The Harvester Press, 1981).
[98] Dawkins, *Selfish Gene.*
[99] P. J. Boyd and R. Richerson, *Not by Genes Alone: How Culture Transformed Human Evolution* (Chicago, IL: University of Chicago Press, 2006).
[100] Gerald Edelman, *Bright Air, Brilliant Fire* (Penguin, 1992).
[101] Henry Plotkin, *Darwin Machines and the Nature of Knowledge* (Harvard University Press, 1994), 61.

parts of the brain and how they are organised.[102] This is the distinction between the molecular neurosciences (which may just be called neuroscience), and cognitive and behavioural neuroscience.

My approach is of the latter kind, and so I face the challenge of multilevel causation. In *Biological Clocks: Explaining with Models of Mechanisms,* Craver and Robins show how the integration of multiple levels may account for a complex phenomenon, using the example the operations of circadian rhythm. Priority in this example is given to denoting what encompasses a *mechanistic account* of a phenomenon's operation that occurs at multiple levels. The first part of this is distinguishing the type of explanation with which one is concerned. In this case, it is constitutive mechanistic explanations:

> *Consititutive* mechanistic explanations… situate an item with respect to the causal structure internal to the phenomenon to be explained. They describe underlying mechanism (they explain the behaviour of the whole in terms of the organized behaviours of its parts).[103]

The second part involves outlining what is required to bridge levels of explanation under this mechanistic account. The first test is one of relevance to the phenomena being explained, as 'anything less is in some sense incomplete. Anything more adds nothing to the explanation'.[104] Typically a model is never fully complete. Those with more obvious gaps are referred to as *mechanism sketches*, and will normally have *filler terms* to mask unknown aspects. The second test is to ensure that any proposed causation is actually a causal relation, not merely a temporal sequence (a crowing rooster does not cause the sun to rise). The third most significant test is demonstrating a clear *relationship of manipulability*:

> Let X and Y be two variables wholes values represent potential inputs into and outputs from one activity. For example, X could be the variable representing whether a gene is expressed, and Y could represent the concentration of a given

[102] W. Bechtel, 'Molecules, Systems, and Behaviour: Another View of Memory Consolidation', in Bickle, *Oxford Handbook*, 14.
[103] C. F. Craver and K. S. Robins, 'Biological Clocks: Explaining with Models of Mechanisms', in Bickle, *Oxford Handbook*, 43.
[104] Ibid., 47.

> protein in the nucleus. One should require that it be possible to change the value of the effect variable by changing the value of the cause variable. Somewhat more rigorously, variable C is causally relevant to variable Y if and only if one can possibly change the value of Y through an intervention, I, that changes the value of X. The difference between activities and pseudo-activities (whose that do and do not fulfil the norms of explanation already sketched) can thus be diagnosed with tests of manipulability. To rule out possible confounds, one must require, in addition, that X changes Y only via the influence of intervening on C. In particular, one must rule out the possibilities that I changes Y directly, that I changes a causal intermediate between X and Y, or that I, is correlated with some other cause of Y.[105]

Craver and Robins point out there is a difference between a mechanistic explanation of a phenomenon, and that of levels of explanation and *realisation* found and debated in metaphysics. This distinction is relevant to those criticisms of reductionism and causation discussed earlier:

> Levels of mechanisms should not be confused with levels of theory or levels of science, because a single theory in neuroscience often describes multiple levels of mechanism and because single fields of neuroscience often do research that spans multiple levels of mechanisms. Nor should levels of mechanism be confused with levels of size (which need not involve a part-whole relationship, as levels of mechanisms do by definition) or levels of causes (because levels of mechanisms, as apart-whole relation, cannot interact causally). Nor should levels of mechanisms be fused with levels of realisation. In level of realisation one compares two distinct properties of one and the same object…Failure to distinguish levels of mechanisms from levels of realisation has led some philosophers and scientists to worry that higher level phenomena in neuroscience and psychology are causally superfluous.[106]

This worry concerns the flow of causation between levels (a metaphysical problem that the authors hold irrelevant to the defined mechanistic account which primarily concerned with

[105] Ibid., 52.
[106] Ibid., 60.

determining the components of a mechanism that produce a phenomenon (e.g. our circadian rhythms accounting for our associated sleeping patterns)). The argument is the one I have already made regarding the definition and use of reductionism. We can scale the levels of a mechanism to explain particular phenomena, and this is distinct from metaphysical issues of causation to do with realisation.

This understanding does not solve the conceptual *v.* empirical debate, but we can see why Churchland, Searle and Dennett argue for their infallible method so strongly; and we can appreciate that their descriptions of a mechanism, while conceptually incomplete, may be good enough to aid our environmental ethics when it comes to policy directives with real behavioural outcomes.

Chapter 3: Evolution and the Genome

> *A creature equipped with all-purpose intelligence simply cannot invent or be taught moral judgment, cannot be taught how to turn a dislike into a disapproval, because 'getting it' requires a certain kind of brain: a brain with specific kinds of mechanism that are geared for such learning—mechanisms forged from a certain kind of evolutionary process. Far from being a just so story, this evolutionary hypothesis appears to be the best story we have.*[107]

Understanding evolutionary theory and the mechanism of natural selection is essential to understanding any proposed biological account of the origin or origins of morality. In this chapter, I review the basic components of evolution by natural selection. The evolution of morality is explored in chapter four. Having done this we have the materials to look with some clarity at the role evolutionary psychology, cultural evolution and cognitive neuroscience might play in a complete account.

Modern biological accounts of morality have often been criticised for their use of evolutionary theory.[108] As such it is important to clearly grasp this underlying naturalist theory if any such account is to be taken seriously. What is clear from the start of this thesis or any like it is that if we are going to use science in philosophy then *we had better convey the science correctly.* Since many extensions of evolutionary theory are speculative they must be interrogated philosophically (in this way, as the Canberra Planners say, constructive naturalists are constrained by their method). It is important that we recognise and understand certain misconceptions of evolutionary theory, particularly as such misconceptions may impact on our understanding of the origins of morality.

3.1 Charles Darwin's Theory of Evolution by Natural Selection

> *The Darwinian evolutionary paradigm has not only revolutionised the biological sciences, having become the explicative epistemological background for all biological phenomena, but it has also had obvious consequences on ethical and social constructs… In moral philosophy the origin between evolutionary and ecological thinking set off a major revolution in ethics: the recognition*

[107] Richard Joyce, *Evolution of Morality*, 139.
[108] E.g., K. N. Laland and G. R Brown, *Sense and Nonsense* (Oxford University Press: Oxford, 2002).

of an ontological continuity between humans and nature and the concern for an ethics capable of integrating this new perception of humanist place in the world.[109]

Darwin's work operates as the *crème du lux* of various kinds of reverse engineering theorising concerning the deep historical roots of morality. The roots of evolutionary theory are found in antiquity and later, in a way much closer to Darwin, in the strikingly acute reflections of his grandfather Erasmus Darwin in the 1790s.[110] When it comes to ethics and understanding the nature of our value systems, the theory has had, as elsewhere, a revolutionary impact.

So what is the theory? It is that when you have organisms subject to variability, differential reproduction and heritability you have evolution by natural selection.[111] It is also accepted that there must be more to the story. For example, we have no quantitative predictions from the theory as far as the number of generations needed for the evolution of behaviours like ours to have achieved the complexity it has.[112] (This problem has left the door open to claims like that of Leslie Valiant, who has suggested this gap will be closed by computer science in the form of biological algorithms (ecorithms)). In a more articulated form, Darwin's theory can, as James Lennox shows us, be specified in nine points. Points 1 to 6 lay out the 'facts' of nature that Darwin studied, while Points 7 to 9 lay out Darwin's theoretical inferences:

1. Species are comprised of individuals that vary ever so slightly from each other with respect to their many traits.
2. Species have a tendency to increase in size over generations at an exponential rate.
3. This tendency, given limited resources, disease, predation, and so on, creates a constant condition of struggle for survival among the members of a species.
4. Some individuals will have variations that give them a slight advantage in this struggle, variations that allow more efficient or better access to resources, greater resistance to disease, greater success at avoiding predation, and so on.
5. These individuals will tend to survive better and leave more offspring.
6. Offspring tend to inherit the variations of their parents.

[109] Ibid., 7.
[110] Leslie Valiant, *Probably Proximately Correct: Natures Algorithms for Learning and Prospering in a Complex World* (New York: Basic Books, 2013), 15.
[111] Darwin, *On the Origin of Species.*
[112] Valiant, *Probably Proximately Correct*, 17.

7. Therefore favourable variations will tend to be passed on more frequently than others, a tendency Darwin labelled 'Natural Selection'.
8. Over time, especially in a slowly changing environment, this process will cause the character of species to change.
9. Given a long enough period of time, the descendant populations of an ancestor species will differ enough to be classified as different species, a process capable of indefinite iteration. There are, in addition, forces that encourage divergence among descendant populations, and the elimination of intermediate varieties.[113]

Despite Niles Eldridge and Gould downplaying the importance of point 8, favouring instead stability as the norm and the paleolithic record as one of punctuated equilibrium, Lennox's account is widely accepted, and the one we shall move forward with. As we can see**,** the fundamental elements of Darwin's theory are variation, inheritance and selection. That variation that leaves the most offspring given the environmental conditions is 'selected'. For example, being born with webbed feet in a time of flooding might lead to a survivalist advantage – though this fact, if it is one, does not mean under any interpretation that the variation, the webbed feet, occurred in response to the environmental condition (flooding).

The surviving offspring with their environmentally 'selected' traits increase in frequency over time (given such traits are still fitness enhancing) until they become species typical. This model of evolutionary theory is now known as the 'chance-variation and selective-retention model'. In *The Structure of Evolutionary Theory*, renowned evolutionary biologist Stephen Jay Gould revived an illuminating metaphor to that helps in understanding this model.

> Selection works like a sieve laden with all the individuals of one generation. Surrounding environments shake the sieve, and particles of a certain size become concentrated, while others pass through the webbing… Sieving represents the causal act of selection—the interaction of the environment (shaking the sieve)

[113] James Lennox, 'Darwinism', in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford University, Fall 2010), http://plato.stanford.edu/archives/fall2010/entries/darwinism/.

> with varying individuals of a population (particles on the sieve)… survival depends causally upon variation in emergent properties of the particles. [114]

Alternatively, one might suggest that webbed feet evolved in response to the environmental flooding, and indeed this is the basis of the evolutionary theory of the French naturalist Jean Baptiste Lamarck. It is relevant to consider briefly as the correct use of evolutionary theory is paramount to its later use with regards to ethics.

> On the Lamarckian view, variations arise in an organism as a direct response to environmental stress or demand, giving rise to a stimulus, which in turn elicits a physiological response, which finally can be passed on via reproduction to offspring. Variations are not chance or random, since they are an appropriate response to an environmental stress. Here 'chance' signals a lack of relation or connection to adaptive needs, an idea akin to, but ontologically quite distinct from, the contrast between 'chance' and 'design'. [115]

A Lamarckian variation is *not* a Darwinian adaptation:

> An adaptation is some form of organization of the phenotype relative to some feature of environmental order. Every adaptation has this dual characteristic of organismic organization and environmental order. It is precisely this relational quality of adaptations that gives them the appearance of being goal or end directed.[116]

Since Fisher (1930) and Haldane (1927) demonstrated that Lamarckian inheritance is not required to explain diversity in the natural world, Lamarckianism has rarely surfaced in scientific enquiry. However, it has not been entirely dismissed, and there has been a recent reconsideration of Lamarckian theory in microbiology. More recently, some have argued that if Lamarck was correct the empirical outcomes would still be the same.[117]

---

[114] Gould, *Structure*, 621.
[115] Ibid., 596.
[116] Plotkin, *Darwin Machines*, 51.
[117] E. Jablonka and M. Lamb, *Evolution in Four Dimensions: Genetic, Epigentic, Behavioural, and Symbolic Variation in the History of Life* (Cambridge MA: MIT Press, 2005).

A further significant point concerning evolution by natural selection is that natural selection so described does not account for the full suite of evolutionary occurrences. Alternatives include genetic mutation, founder effects and genetic bottlenecks.[118] That this is true immediately raises a question concerning our moral and ethical traits and capacities—did they evolve by natural section or are they, for example, the product of genetic mutation? The difference is important in terms of timing, and of so of origins as opposed to *origin*—for did nature build the moral brain from a piecemeal set of environments, and so in a piecemeal way, or did that brain emerge under a certain set of environmental conditions and so, as it were, 'all at once'? We will return to this question later, but it gives us one reason for being cautious about how we might use evolutionary theory to explain morality. James gives four dangerous missteps to be avoided in using evolutionary theory when thinking about the origin(s) of morality, all of which I seek to avoid in this enquiry:

> (1) Conflating adaptation and adaptiveness; (2) conflating explanation and justification; (3) misunderstanding the scope of an evolutionary explanation; and (4) succumbing to the temptation of genetic determinism.[119]

After the 1950s, and in an apparent triumph of reductionism, Darwin's original theory, which focused on selection at the level of the individual organism, tended to be pitched now at the level of genes (i.e. Dawkins' selfish gene).[120] On this view organisms are gene constructed vehicles for gene replication.[121] 'Genetic information is stored in sequences of DNA base pairs, genetic information is expressed as proteins, and ultimately physical structures such as limbs and eyes'.[122] Such prominent scientists as George C. Williams,[123,] Dawkins,[124] and Gould,[125,] all agree that genes replicate, however the idea that they are the sole, even most important, level of selection is no longer so universally held, with Gould and others pointing to multiple levels of selection of (in varying conditions) varying

[118] Scott M. James, *Introduction*, 15.
[119] Ibid., 21.
[120] Dawkins, *Selfish Gene*, 21.
[121] Mendel, 1865.
[122] A. Mesoudi, *How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences* (Chicago, IL: University of Chicago Press, 2011), 3.
[123] George C. Williams, *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton, NJ: Princeton University Press, 1966), 254.
[124] Dawkins, *Selfish Gene*, 21.
[125] Gould, *Structure*, 683.

importance (e.g., cell lineages, organisms, species and clades). Darwin insisted very strongly upon a single-level theory with 'the struggle among 'individual bodies' as a virtually exclusive locus of causality'.[126] However, he did consider the possibility of a causal origin for this, similar to the gene, in his 'Gemmules' concept.[127]

3.2 The Hull Metaphors and Dawkins' Selfish Genes

In 1980, the biologist David Hull introduced two significant metaphorical terms ('replicator' and 'interactor') to conceptually clarify the essential elements of Darwinism.[128] Hull defined a replicator as an entity that passes on its structure directly in replication, and an interactor as 'an entity that directly interacts as a cohesive whole with its environment in such a way that replication is differential'.[129] Hull defined selection with reference to both attributes as 'a process in which the differential extinction and proliferation of the interactors cause the differential perpetuation of the replicators that produced them'.[130] In effect this means that genes are not units of selection, but are bookkeeping devices recording the history of interactors. This is something Dawkins continues to reject. For him 'replicators' must be *the* unit of selection. His logic is presented by Gould in the following three steps:

> a) The unit of selection must be a replicator;
> b) Replicators must transmit faithful or minimally altered copies of themselves across generations;
> c) Sexual organisms disaggregate across generations and therefore cannot be units of selection, but genes qualify by faithful replication.'[131]

In Dawkins own words (according to Gould. 'the purplest prose' in all evolutionary writing):

> Replicators began not merely to exist, but to construct for themselves containers, vehicles for their continued existence. The replicators which survived were the ones which built survival machines for themselves to live in… Survival machines

[126] Ibid., 596.
[127] Darwin, *On the Origin*.
[128] Hull, 'Units of Evolution'.
[129] Gould, *Structure*, 615.
[130] Hull, 'Units of Evolution'.
[131] Gould, *Structure*, 617.

> got bigger and more elaborate, and the process was cumulative and progressive… Four thousand million years on, what was to be the fate of the ancient replicators? They did not die out, for they are past masters of the survival arts. But do not look for them floating loose in the sea; they gave up that cavalier freedom long ago. Now they swarm in huge colonies, safe inside gigantic lumbering robots, sealed off from the outside world, communicating with it in tortuous indirect routes, manipulating it by remote control. They are in you and me; they created us, body and mind; and their preservation is the ultimate rationale for our existence. They have come a long way those replicators. They now go by the name of genes, and we are their survival machines.[132]

Soon after, fellow biologists raced to support Dawkins' theory of inherent selfishness with their own witty remarks: 'Scratch an altruist … and watch a hypocrite bleed'.[133] Adding further weight to his argument, Dawkins insisted genes proceeded organisms in time.[134]

Gould was having none of this. He saw the selfish gene theory as 'a confusion of book keeping with causality'.[135] He argued instead that 'the causality of selection resides in interaction, not in replication'.[136] He suggested a fully hierarchical theory of selection, that is, 'a full genealogical hierarchy of inclusion—with rising levels of genes, cell lineages, organisms, demes, species and clades-features clearly definable validating… an extension and reformulation of Darwin's exclusively organismal theory into a fully hierarchically theory of selection'.[137] Gould argued that 'selection at one level may enhance, counteract, or just be orthogonal to selection at any adjacent level. All modes of interaction prevail among levels and make prominent imprints in nature',[138] and so there may be 'any number of units' of selection.[139]

In 1970, Richard Lewontin found that genes do not behave as individual units during reproduction, which thus disqualifies them from meeting the definition of a replicator. In

[132] Dawkins, *Selfish Gene*, 21.
[133] M. T. Ghiselin, *The Economy of Nature and the Evolution of Sex* (Berkeley, CA: University of California Press, 1974), 274.
[134] Ibid., 36.
[135] Gould, *Structure*, 614.
[136] Ibid., 615.
[137] Ibid., 72.
[138] Ibid., 677.
[139] Ibid., 682.

1985, Susan Oyama demonstrated that in phylogeny and ontolgeny genes are not the primary causal factors in the relation with the environment, and dismissed gene-centred view of evolutionary theory on this ground.[140]

3.3 Group Selection and Altruistic Behaviour

Prior to Dawkins' publication of *The Selfish Gene,* the dominant explanation for altruistic behaviour was the notion of group selection, which, according to Frans De Waal, was actually recognised by Darwin himself in his final work, *The Descent of Man*:

> It's fine to describe animals (and humans) as the product of evolutionary forces that promote self-interests so long as one realizes that this by no means precludes the evolution of altruistic and sympathetic tendencies. Darwin fully recognised this, explaining the evolution of these tendencies by group selection, instead of the individual and kind selection favoured by modern theoreticians.[141]

Others disagree with this reading, seeing Darwin's remarks as critical speculation rather than the acknowledgement of a potential flaw in evolutionary theory.[142] The conflict between group and individual selection with regards the explanation of altruism was made famous in a debate between George Williams (1966), a proponent of individual selection, and Wynne-Edwards (1962), who argued for the existence of group selection. A major intervention was that W. D Hamilton (1964) who argued that altruism could be accounted for by individual selection. His rule is simply $c<rb$, where, c represents the cost to the actor, b demonstrates the benefit to the recipient, and r illustrates the degree of relatedness between actor and recipient. It follows if the cumulative benefit, given sufficient next of kin proximity, is greater than the cost of the action to the actor, then it benefits the actor to behave in such a manner. This is a theory of total genetic benefit, otherwise referred to as *inclusive fitness*:

[140] S. Oyama, *The Ontogeny of Information: Developmental Systems and Evolution* (Cambridge, England, 1985), 197.
[141] Frans de Waal, ed., *Primates & Philosophers: How Morality Evolved* (Princeton, NJ: Princeton University Press, 2006), 14.
[142] S. Viren, *Evolutionary Psychology* (West Sussex: John Wiley & Sons, 2011), 89. (But see Darwin, *On the Origin*, 236.)

> According to Hamilton's rule, with a coefficient of relatedness this high, natural selection will strongly favour behaviour that benefits sisters, even at the cost of not reproducing themselves. Adopting the gene's-eye point of view, then, should bring into focus the root of many kinds of helping behaviour among non-human animals.[143]

Additionally, there was felt to be another problem for group selection as an explanation for altruism: a group of helpers is almost always vulnerable to what Dawkins called 'subversion from within', for *within* groups there is reason to think that selfishness will tend to drive out or trump altruism. For 30 years individual selection was presumed correct and group selection unnecessary, but more recent work by David Wilson and others suggests otherwise.

The idea is that while the evolution of group-level functional organisation (coordinated example of behavioural altruism) cannot be explained on the basis of natural selection operating within groups, it may be explained in terms of natural selection between groups. Wilson argues that altruism evolves when between-group selection prevails over within-group selection:[144] 'Selfishness beats altruism within groups. Altruistic groups beat selfish groups. Everything else is commentary'.[145]

Today there is less of an issue with concluding that 'group selection could be an additional level above the gene-level, where the fitness of whole groups is subject to selection pressures'.[146] Indeed, some proponents of evolutionary psychology explain morality in a limited/narrow sense along these lines.[147] Moreover, the basic trade-offs that create conflicts between levels of selection do not depend upon the mode of inheritance or the distinction between directed *v.* undirected variation.[148] This debate is not settled however, as some like Churchland, continue to argue that there is no need for group selection whatsoever.[149]

[143] Scott M. James, *Introduction*, 34.
[144] David Sloan Wilson, *Does Altruism Exist*?, 21.
[145] Ibid., 23.
[146] Looren de Jong, 'Evolutionary Psychology and Morality', 110.
[147] Ibid., 118.
[148] David Sloan Wilson, *Does Altruism Exist?*, 57.
[149] Churchland, 'Inference', 423.

3.4 Genetic Influence: Back and Forth

> *Brains are built by genes. Unless the genes build a brain that is organized to avoid danger and seek food, water, and mates, the animal will not long survive nor likely reproduce. If you have no offspring, your genes do not get passed on. So if you had genes that built your brain so that you had no regard for your well-being, those genes would expire with the brain they built.*[150]

On Churchland's view the significance of genes as causally underlying the nature of our experience should not be overlooked simply because certain biologists have overstated the case. The notion that the brain and mind simply supersede, in causal terms, that of the gene is itself, is a limited view given that *genes build brains*.

A specific gene might create or alter serotonin levels in a specific neural network, which in turn causes that individual to behave in certain ways given certain environmental cues. Basic gene-to-behaviour causation can be demonstrated with two genetically identical mice: remove a certain gene in one of the mice and a behavioural difference occurs; this is a deterministic biological fact, as repeatable and as consistent as measuring in experiment the effect of gravity on an apple.[151] As humans are more complex any causal gene-to-behaviour theory will have to navigate this complexity.

This complexity is pointed to by Tim Spector, who has spent his life studying genes and twins. Cloned cats are not true to the original in coat colour because that trait is an unpredictable *epigenetic* phenomenon. Thus, removing mice genes may have clear causal outcomes in terms of some behaviour, but this does not necessarily mean that having specific genes will result in specific traits, as there is now emerging an epigenetic level to consider as well:

> It may be more accurate to think of genes as defining a space of potentialities for an organism: the range of phenotypes that could result under different environments and different background conditions.[152]

---

[150] Patricia S. Churchland, *Touching a Nerve: The Self as Brain* (New York: Norton and Company, 2013), 85.
[151] B. A. Dyck et al., 'Behavioural Abnormalities in Synapsin 11 Knockout Mice Implicate a Causal Factor in Schizophrenia'. *Synapse* 63, no. 8: 662–672.
[152] A. L. Roskies, 'What's "NEU" in Neuroethics?' in Bickle, *Oxford Handbook*, 455.

The simple idea that genes are directly responsible for the breadth of human behaviour is insufficient when we consider the sheer complexity of a growing human and the environmental conditions with which it interacts. The diverse capacities of the brain, understood at a functional level of analysis, seems a more promising approach than any focus on gene causation alone.[153] The simplistic notion that genes might control or be the source of moral judgement—two different propositions—have largely been left behind, though occasionally the case is still made.[154] For example, a relatively recent study argued for specific genes that account for the capacity to punish others,[155] and another set for antisocial personalities.[156]

Gene development is essentially the result of switching on and off structural genes at different times by a diversity of regulatory genes, and it is the difference in regulatory genes that result in obvious differentiation in organismal body plans and behavioural capacities.[157] In the case of early developmental genes, alteration has been shown to remove certain behaviours (e.g. nurturance behaviour in mice).[158] According to Churchland, we are beginning to understand the relationship between gene and epigenetic factors, and their impact on what we take to be socially acceptable behaviour, which in some contexts *seems* to define the very nature of the particular moral spectrums we have. Churchland, referring to a study of rats, points out that the motherly act of being licked results in a chain of neurochemical events within the offspring, which directly alters gene expression and in turn modifies social capacities.[159]

To better understand the causal impact of genes we first need a brief understanding of what exactly they are, and what they can and cannot do. While we understand that their causal power is limited, so not eliminative, that influence must still be understood. So what are genomes, chromosomes and DNA? It is best to turn to the experts to define scientific

[153] Churchland, *Touching a Nerve*, 87.
[154] See e.g. A. Raine, 'From Genes to Brain, to Antisocial Behaviour', *Current Directions in Psychological Science* 17, no. 5 (2008): 323–328.
[155] McDermott, R., D. Tingley, and T. J. Cowden, 'Monoamine Oxidase A Gene (MAOA) Predicts Behavioural Aggression Following Provocation', *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 7 (2009): 2118 –2123.
[156] M. DeLisi et al., 'All in the Family', *Criminal Justice and Behaviour* 36, no. 11 (2009): 1187–1197.
[157] A. Rosenberg, 'Lessons for Cognitive Science from Neurogenomics', in Bickle, *Oxford Handbook*, 145.
[158] Ibid., 146.
[159] Churchland, 'Inference', 420.

subject matter rather than relay it in a manner that may suggest or encourage false causal inferences. First let's look at what a *gene* is made up of.

> Each chromosome consists of two very long thin strands of DNA chains twisted into the shape of a double helix and are located in the nucleus (the 'control centre') of our body cells. These can be thought of as long strips of genes. Genes are also located in very small compartments called mitochondria that are randomly scattered in the cytoplasm of the cell outside the nucleus. All of the DNA in the cell (in the nucleus and the mitochondria) make up the genome.[160]

The *genetic code* is essentially a chemical code that gives certain instructions for biological production. The code is made up of very long interlinked information but it only has four source components: Adenine (A), Guanine (G), Thymine (T) and Cytosine (C). Each gene is comprised of three of these letters as they tell the cell to produce a particular amino acid at a particular time. It is the *sequence* in the gene that enables the cells to assemble the amino acids in the correct order to make up a protein.[161] Thus our bodies and brains are made up of cells, and genes can control the behaviour of these cells. This is referred to as *gene expression*, the causal reaction of which defines bodily form and its growth, subject to environmental influence.

In all this, modern genetics is replacing the old idea that genes were instructions to build offspring. Rather, they are a functioning part of the brain and body within our (the organism's) life.

It is now recognised that genes do not act independently; they require assistance in synthesising a protein by an epigenome.[162] This tells against strict genetic causation, for it breaks down the linear DNA to protein causation. Yet we must also accept that altering gene expression can alter the observable traits of an organism.[163] Recent experiments such as Kosfeld's have shown genetic modification may change our behaviour. Kosfeld administered oxytocin to control gene expression and to influence an investor's level of

[160] Barlow-Stewart, 'Genes and Chromosomes: The Genome', *NSW Government Health Centre for Genetics Education*, last modified 2012, http://www.genetics.edu.au.
[161] Ibid.
[162] E. Charney and W. English, 'Candidate Genes and Political Behaviours', *American Political Science Review* (2012): 7–10.
[163] Churchland, *Braintrust*, 100.

trust.[164] Similarly, but perhaps more obviously, Heike Tost (and others) have suggested that carrying specific types of genes results in specific types of behavioural responses in social situations: A famous example concerns the rs5357A allele, which is known for generating less positive or negative feelings about social interactions.[165] The field of developmental biology has even managed to map and program the construction of basic life forms:

> In the case of Drosophila, molecular biologists have been able to identify a literal program of development (one that can be implemented by a laptop computer) involving some 30 genes and their gene products and by the repeated implementation of a structure of subprograms that builds the Drosophila embryo.[166]

The causal link seems strong, yet again the issue of complexity arises, and so this kind of evidence has been dismissed by many on the grounds of causal dissipation.[167] Even gene causation scientists like Tim Spector now see the gene-to-behaviour argument as limited. Rosenberg too is highly sceptical of genes-to-behaviour causation, arguing for the environment as having a greater influence than genes.[168]

Another way to approach this issue from is the study of newborn behaviours. Paul Bloom in 'Just Babies: The Origins of Good and Evil', seems to have shown that there is an innate, so genetic, capacity for punishment. He found that '1-year-olds who feel that a puppet that doesn't play nice with other puppets should be punished—in fact, they'll personally do the punishing, by taking the puppet's treats away'.[169] If we accept this evidence then we might say we are equipped to perform certain types of moral behaviours thanks to our initial genetic construct and this equipment has a defined spectrum. But surely we need to be more specific than simply a 'gene *for* cooperation'?[170] Babies may come with some

[164] Ibid., 72.
[165] Ibid., 53.
[166] Rosenberg, 'Lessons for Cognitive Science', 145.
[167] J. Flint, R. Greenspan, and K. Kendler, *How Genes Influence Behaviour* (New York, NY: Oxford University Press, 2010), 211.
[168] Ibid., 188.
[169] Robert Wright, 'Why Can't We All Just Get Along? The Uncertain Biological Basis of Morality', *Atlantic*, 23 October 2013. http://www.theatlantic.com/magazine/archive/2013/11/why-we-fightand-can-we-stop/309525/.
[170] Rosenberg, 'Will Genomics Do More?', 187.

general genetically determined boundaries for moral behaviour, but we certainly cannot say that every action performed by a child has a gene specifically for it, a point *neuroscientists* reinforce over and again:

> Keep in mind that genes make proteins, and that there are very long causal routes between proteins and brain circuitry, and further long causal routes between brain circuitry and environment, which in turn affects gene expression and proteins. Genes are part of genetic networks, and these networks interact with the environment in complex ways… nothing very specific can be concluded, such as that there is a 'gene for' fairness, or religion, or wanderlust.[171]

The alternative put forward by Churchland and others points to general learning purposes within the brain's structure, which account for our 'plasticity of mind'.[172] (So, for instance, writing was invented without a new gene for it.) In the context of genes Churchland goes on to dismiss Tost's casual findings because of the ubiquitousness of pleiotropy.

> *Pleiotropy*—when a gene plays a role in many different, and functionally distinct, aspects of the phenotype (traits the organism has)—turns out to be not the exception, but the rule…. Consider the physical trait of height in humans is associated with 54 known alleles, but collectively, they account only for 5% of the heritability of height. The rest is a mystery.[173]

Neuroscience is more promising here, for it incorporates environmental change including the effect of culture more directly. After all, altering the brain's neural networks during a lifetime is, biologically, far easier than altering a gene.[174]

Another different way of viewing genes is to see them as *constructive rules*[175] that define the limits of this available behavioural spectrum. But how can a *gene* be a *rule*? Where is the normativity? And if it is defined in terms of its 'proper' contribution to the organism, then as a rule it is derived from the organism, not from the gene. We can say, for example,

---

[171] Churchland, *Braintrust*, 22.
[172] Ibid., 20.
[173] Ibid., 98.
[174] Churchland, *Touching a Nerve*, 87.
[175] E. O. Wilson, *Consilience: The Unity of Knowledge* (New York: Random House, 1998), 138.

because of our evolved genes we can 'cooperate', but we must also admit that the genetic story ends at a point where modern neuroscience and cultural information continues on ('cooperation via sharing the culturally valued vegemite and toast') and that further ahead lies the realm of conscious reasoning.

In summary, in the evolutionary approach genes provide the first level of the causal biological story. For the needed further levels we need to look at neuroscience:

> Genetics and genomics have provided fertile ground for many ethical reflections on human nature, but the relationship between the brain and the self is far more direct than the link between genes and personal identity. The locus for integrating behavior resides in the brain, even if our genes determine discrete features. Whether neurotechnology measures that behavior through imaging, or manipulates it through implants of neural tissue or devices, it will fundamentally alter the dynamic between personal identity, responsibility and free will in ways that genetics never has.[176]

3.5 An Evolutionary Platform

> *Behavioural dispositions, norms and social institutions are not among the hard parts preserved in the fossil record.*[177]

There is another problem with utilising evolutionary theory in the way I want in this thesis. For if we are *really* honest the evidential material of hominoid evolution is very scarce. Claims about brains developing due to certain X conditions/environments may well be largely guesswork, and sometimes guesses determined by what we want to find (as we reverse engineering evolutionary theory itself in favour of ideas that get us to where we want to be). Even recognising the anthropological tradition (Sterelney's work) Investigator bias then is a very real threat, even if it is based on the latest neuroscience. My attempt to combat this healthy scepticism to evolutionary explanation is to adopt a multileveled (and non-eliminative) naturalist account, and to do so on the basis of a conviction well expressed by Richard Joyce:

[176] J. Illes and E. Racine, 'Imaging or Imagining? A Neuroethics Challenge Informed by Genetics', *American Journal of Bioethics* 5 (2005): 14.
[177] Rosenberg, 'Will Genomics Do More?', 179.

> How in neurological or genetic terms did natural selection bring about the human linguistic faculty, or our preparedness to find snakes frightening (as opposed to guns), or our ability to distinguish between male and female faces, or even something so basic as our interest in sex? The fact that we don't know the complete answer to any of these questions should stimulate our inquiry rather than nourish suspicion that such phenomena are not the direct product of natural selection at all.[178]

This chapter has been an introduction to evolution. With evolutionary theory we have an account (or the framework for an account) that fits the reductive deterministic picture. It provides a general historical, causal, way of approaching and understanding our behavioural capacities and incapacities, dispositions and traits. On one level, our behaviour is a product of a specific biological construct through time. We have, as part of this construction, the capacity to reason. Our brains likely evolved to make decisions and solve problems in *particular* environments. People making decisions are no longer simply our assigned metaphors representing their behaviours as 'ignorant', 'greedy', 'lazy' or 'grossly incompetent', for example, but rather are viewed as biological creatures with a specific suite of capabilities that they utilise in capacities defined through specific histories.

Recognising that there is debate between group selection and individual organism selection, and debate over whether the unit of selection is indeed the gene or not, yields avenues for further exploration from the decision-making lens. We can see that today many hold evolution to operate at multiple levels in terms of comprising the organism, and David Wilson's 2015 work cements this view. Our conclusion is as profound as it was when it was first realised—the natural world operates in a certain way, with a certain causal consistency we can discern through the theory of evolution by natural selection (and other evolutionary occurrences such as mutation). From this platform we can explore the sociobiological fields and the causal claims they make.

[178] Richard Joyce, *Evolution of Morality*, 133.

Chapter 4: The Evolution of Morality

4.1 Defining the Subject

> *Darwin firmly believed his theory capable of accommodating the origins of morality and did not see any conflict between the harshness of the evolutionary process and the gentleness of some of its products. (Darwin 1982 [1871] 71-72)*[179]

Altruism in nature raises an important question in the quest for the origin of morality, for how do we define altruism, and is it, so defined, sufficient for morality? Are ants altruistic in the relevant sense? Or is any 'altruism' we might discern there something different from that we find in human communities? Is altruism an observed behaviour (and defined in these terms), or is it behaviour that is defined only in terms of its antecedent causation? Centrally, of course, as the result of 'moral deliberation and judgement'?[180]

To approach these questions it is useful to start with the precise meaning evolutionary biologists give to cooperation and associated terms:

1. A behaviour is social if it has fitness consequences for both the actor and the recipient.
2. A behaviour that is beneficial to the actor and costly to the recipient (+/-) is selfish.
3. A behaviour beneficial to both is mutually *beneficial* (+/+)
4. A behaviour that is beneficial to the recipient but cost the actor is *altruistic* (-/+)
5. A behaviour that is costly to both actor and recipient is *spiteful* (-/-)
6. Whether a behaviour is costly or beneficial is defined on the basis of:
   (i) the lifetime fitness consequences (not just the short-term consequences)
   (ii) the fitness consequences relative to the whole population, not just relative to the individuals or social group with which the individual acts.

Compare this to what philosophers tend to say about morality and we enter into complex territory:

[179] de Waal, *Primates & Philosophers*, 14.
[180] Scott M. James, *Introduction*, 50.

- The general capacity for normative judgment and guidance, and the tendency to exercise this capacity in social life;
- The capacity for certain sentiments and the ability to detect them in others;
- The tendency to experience and to be motivated by certain sentiments in certain types of situation;
- The tendency to make certain particular kinds of moral judgment or inference, or to have certain characteristic moral intuitions (i.e., a 'moral sense');
- The tendency to exhibit certain particular types of behaviour in certain types of situation;
- The tendency of societies to exhibit certain particular systems of norms or types of practice.[181]

According to philosopher Richard Joyce, one of the key elements to ironing out the potential confusion between *merely* altruistic and cooperative behaviour as opposed to their *moralised* expression lies in the idea of *prohibition*. Only those behaviours (however much they are altruistic in the evolutionary biologist's sense) are morally altruistic which are performed (or deliberated on) in the light of the idea that not to so act is, in some meaningful sense, prohibited. Indeed, for Joyce, this 'prohibition awareness' defines what it is to possess a moral sense.[182] On this view the fact that members of a species may display behaviours that are costly to themselves even as it benefits their fellows does not itself establish that they are behaving morally. In a similar vein, Scott James summarises the key components that seem to distinguish moral creatures from non-moral creatures:[183]

> (1) Moral creatures understand prohibitions.
> (2) Moral prohibitions do not appear to depend on our desires, nor;
> (3) do they appear to depend on human conventions, like the law. Instead, they appear to be objective, not subjective.
> (4) Moral judgments are tightly linked to motivation: sincerely judging that some act is wrong appears to entail at least some desire to refrain from performing that act.

[181] Richard Joyce, *Evolution of Morality*, 133.
[182] Ibid., 51.
[183] Scott M. James, *Introduction*, 56.

(5) Moral judgments imply notions of desert: doing what you know to be morally prohibited implies that punishment would be justified.
(6) Moral creatures, such as ourselves, experience a distinctive affective response to our own wrongdoing, and this response often prompts us to make amends for the wrong doing.

Particularly important here is the final condition – for it connects together the idea of a prohibition with an internal self-prohibition in a way that lets us start to see how conditions 2-5 might be filled out. It is such self-imposed prohibitions (rather than merely other-imposed prohibitions, as when a creature encounters the violently enforced territorial boundaries of another creature) that – as guilt and/or shame – bind us into morality, and into morality as something of unique authority and motivational potency.

In fact most current literature coming from the biological sciences ignores this internal conception of self-prohibition, but perhaps this is not as fatal or serious a problem as it may seem. For it would seem undeniable that internal self-prohibition arises from and through a process of external prohibitions, however the former derives from the latter. In this case it seems plausible that a different, if related, notion – that of *cooperation* – has a certain priority. This is because the very move from purely external to self-imposed prohibitions would seem to require a framework of communal actions which together (co-operatively, in terms of blame and responsibility) constructs or produces this altruistic enabling self-regulation.

Cooperation can be defined as a behaviour that provides a benefit to another individual (recipient) and whose evolution is dependent on the beneficial effect for the recipient.[184] The easiest way for evolution to facilitate moralised cooperation (a cooperation that may arise, at its origin, from facts the facts of species sexual reproduction strategies which, for 'social species' like humanity with its long period of offspring dependence, presuppose functional cooperation) is likely to have been a matter of developing a set of moral rules utilising various emotional states. We might think of such emotions as navigation markers,

[184] Churchland, *Braintrust*, 66. Churchland summarises definitions from S. A. West, C. El Mouden, and A. Gardner, '16 Common Misconceptions about the Evolution of Cooperation in Humans', *Evolution and Human Behaviour* (in press), http://qqq.zoo.ox.ac.uk/group/west/pdf/west_etal.pdf).

less complicated than the evolutionary development of vision through the eye and brain, but performing a similar function in terms of increasing the chance(s) of survival.

What begins to emerge from this brief evolutionary account is a practical opportunity in terms of a decision-making and policy construction, based on a piggybacking of existing evolutionary functionalities.

4.2 A Necessity to Cooperate

I have given a brief definition of cooperation from a behavioural level, and now I want to unpack cooperation further in light of the pragmatic opportunity I have pointed to. To do this I have integrated Scott James' simple story of the evolution of morality with Dan Sperber, Nicolas Baumard and Jean-Batiste Andre's 'mutualistic approach to the evolution of morality' in the following three points:[185]

1. Natural selection creates a psychological facet (foods look delicious) to achieve the biological purpose (survival). What we know is the food looks delicious, what we do not have knowledge of is why we need it or how our bodies utilise it. The evolution of cooperation is no different; we feel moral inclinations, indeed certain emotions towards others, because cooperative behaviour allowed for survival.
2. Individuals were in competition to be involved in cooperative tribes. Morality then, as the suite of features described, arose as an adaptation, for it was best to treat others with impartiality and to share the costs and benefits of cooperation. To offer less or more in the cooperative effort would result in being rejected or exploited.
3. The moral adaption was not a 'flick of the switch'. While evolution may have designed individuals to establish and preserve cooperative alliances by acting impartiality this was occurring in the face of existing selfish capacities. Such existing hardware was in time overcome by an evolved suite of moral safeguards, including retribution, punishment, guilt, and generosity.

As discussed in 3.3, David Wilson takes these cooperative abilities to have evolved from the prominence of group level selection, rather than individual level selection.[186] What is

[185] Dan Sperber, Nicolas Baumard and Jean-Batiste André, 'A Mutualistic Approach to Morality: The Evolution of Fairness by Partner Choice', *Behavioural and Brain Sciences* 36 (2013): 59–112.
[186] David Sloan Wilson, *Does Altruism Exist?*, 49.

controversial about this view is that it suggests the 'moral suite' arose as a kind of 'flick of a switch', in which a single shift in the balance between levels of selection resulted in an 'entire package of distinctively human traits, including our ability to cooperate in groups of unrelated individuals, our distinctive cognition, and our ability to transmit culture.'[187] This idea directly challenges the Scott-Sperber-Baumard position.

The matter here is far from settled argument, and a real issue remains with showing how the 'moral suite' develops, and what goes into that suite. For how does one define the evolution a specific moral ingredient, such as patience? Indeed why is (or is) patience a moral ingredient at all? A plausible answer is that it enables us to fend off the temptation to cheat and hang on to our reciprocal relationships.[188]

Plausible or not, we seem to be engaged in a reverse engineering guessing game that can only be properly constrained if we can spell out the evolution of the 'moral suite' without circularity and ambiguity. As Hauser points out, with punishment in mind:

> We don't know when this capacity evolved in our species, we also don't understand which aspect of the environment changed, creating pressure for punishment to evolve… whatever happened, and whenever it happened, the landscape for cooperation changed.[189]

Is punishment (or patience) crucial to the evolutionary story of morality from and through cooperation? Perhaps punishing each other ensured collaboration could be resumed despite our selfish genetic hardware? Certainly the punishment/prohibition dialectic seems to be an evolutionary capacity that is *almost* solely human.[190]

A potential problem with any evolutionary account of morality as a system of laws or rules is that in fact morality differs across cultures. This may be used as evidence that morality is not itself innate, or that what are innate are certain general learning mechanisms that account for this kind of thought and behaviour (Section 4.4 will cover nativism issues). If

[187] Ibid., 52.
[188] M. D. Hauser, *Moral Minds*, 448.
[189] Ibid., 448.
[190] Ibid., 443.

the chosen way is the latter, then often 'reason' is seen as the foundation for morality proper, not the emotions.

What is clear is that as yet we have no scientific consensus on the origin of (our) moral sensibility (we may even contend about whether morality arose due to a mutation or via natural selection), however I want to insist – and reinforce the point – that the pragmatic opportunity is worth considering with the cooperative lens.

My ultimate concern is with the ethics of sustainability in the Anthropocene. Now some environmentalists have cast sustainability in terms of 'natural capital', and arguing that an ethic of sustainability means 'we must learn to live off the interest, not depleting the capital but holding it in perpetuity'.[191] This seems good Anthropocene sense, but good sense is not enough here. We must, as Rees said, *learn* to do this. And so the real question is how one can go about learning to behave in this way?

It seems plausible that we have done so by evolutionary necessity—cooperate, or perish, and punish those who do not cooperate. We learnt to cooperate with each other and survived because of it, but can we now learn to cooperate together in a way that ensures our sustainability, and especially when this is, in the Anthropocene, a global challenge of a kind never before encountered? In such circumstances – precisely brought about through our traditionally evolved dispositions and behaviours – it would clearly be advantageous if the next phase of our evolution were *proactive*. But, how? Well, surely a first thing to do is to look for constructions derived from our initial cooperative behaviours – at least in the sense of what might work and what pretty clearly won't work, given our natures. For example, can we exploit the truth there is in the following fact?

> A blatant disregard for the environment is linked to a general disregard for public goods, but someone who exhibited no concern for public goods would run the risk of diminishing her reputation as a "socially conscious" person.[192]

4.3 Attachment and Bonding

[191] Rees, 'Sustainable Development', 661.
[192] Scott M. James, *Introduction*, 81.

> *With the evolution of mammals, the rudimentary self-caring organisation was modified to extend the basic values of being alive and well to selected others to me and mine.* [193]

We have already heard the general cooperative story about the evolution of morality. Now I want (in my non-eliminative reductionist approach) to further my search for a story about morality by looking at the brain's development (though leaving the actual neuroscience for latter chapters).

Patricia Churchland is a well-known neuroscientist who started out in philosophy. Based at the University of San Diego, she now practices 'Neurophilosophy', the intersection of neuroscience and philosophy. Her main hypothesis is that morality originates in the *neurobiology* of *attachment* and *bonding***.**

> The hypothesis on offer is that what we humans call ethics or morality is a four-dimensional scheme for social behaviour that is shaped by interlocking brain processes: (1) caring (rooted in attachment to kind and kith and care for their wellbeing), (2) recognition of others psychological states (rooted in the benefits of predicting the behaviour of others), (3) problem solving in a social context (how should we distribute scare goods, settle land disputes; how we should punish miscreants) and, (4) learning social practices (by positive and negative reinforcement, by imitation, by trial and error, by various kinds of conditioning, and by analogy.)[194]

A challenge to the simple picture provided by Churchland comes in the form having to explain *modern* moral elements:

> How do we get from familial caring to broader community-wide values such as honesty, loyalty, and courage? The answer has two intertwined parts: learning by the young and problem solving by everyone. [195]

[193] Churchland, *Touching a Nerve*, 98.
[194] Churchland, *Braintrust*, 9.
[195] Churchland, *Touching a Nerve*, 104.

Churchland's view is that there was an evolutionary reorganisation of neural structures that led to modern-day moral functioning. At the heart of her account, though by no means all of it, is the relevance of the neurochemical Oxytocin:

> Oxytocin, a very ancient peptide (chain of amino acids), is at the hub of the intricate network of mammalian adaptations for caring for others, anchoring the many different versions of sociality that are seen, depending on the evolution of the lineage... Besides the role of oxytocin and other hormones, two additional interdependent evolutionary changes in the brain were crucial for mammalian sociality that prefigured morality… [1] negative feelings of fear and anxiety in the face of separation or threat to the offspring… [2] an increased capacity of learning, linked to pain and pleasure, that served an individual in acquiring detailed knowledge of the 'ways' of others in the group. [196]

Add to this the development of mirror neurons with regard to social interaction and Churchland offers a persuasive picture of how our moral functionality may have evolved, but as the German Philosopher Jan Salby points out (or in Hickock's critique) in his critical review of Churchland's *Brainstrust*,[197] as it stands the story is not enough. To go further in developing Churchland's approach we need to investigate the question of Nativism and the so-called Emotion v. Reason debate.

4.4 Nativism/Innateness

Churchland writes:

> Recently Marc Hauser, a psychologist and animal behaviour scientist, defended the innateness approach to morality. Hauser thinks there are universals in human moral understanding – views about what is right and what is wrong – that obtain in all societies. These universals are he contends invisible in the unreflective intuitions that people summon in addressing a specific moral issue. For example Hauser finds that there is widespread agreement that incest is wrong, and that dinking apple juice from a brand new bedpan is disgusting…. Hauser argues that

[196] Churchland, *Braintrust*, 14.
[197] Slaby, Jan, Crossing the Is/Ought Boundary, *The Hedgehog Review*, 48.

> humans… have a 'moral organ' that specifies the universal principles of morality.[198]

Marc Hauser is not alone in his quest to demonstrate innate moral universals. Steven Pinker, John Mikhail, Noam Chomsky, John Rawls, Gilbert Harman, Frans de Waal, Daniel Dennett and Richard Joyce, amongst others, all support a form of nativism.

In this analysis I look at the evolutionary story (cooperation, bonding etc) in terms of the physical 'mechanics' of the moral sense**.**

Relying on a study in the 1980s by Tisak and Turiel,[199] which showed children as young as three making the distinction between moral rules and conventional rules, Mikhail argues that children must have obtained this capability to so distinguish by an innate sense, for they could not have been trained, nor, by the age of three, have learned the distinction**.** The inference to the best explanation is – as it was for the language faculty generally for Chomsky – that there is an innate moral capacity. Drawing on this commonality Mikhail shows that Chomsky's answer three questions on language, have their parallel when it comes to morality (Table 4.1).[200]

Table 4.1: Questions and Answers on Language and Morality

Language	
Question	Answer
(a) What constitutes knowledge of language?	Generative Grammar: The brain knows natural language 'English or Japanese or Russian' etc.
(b) How is knowledge of language acquired?	(UG) Universal Grammar – A subsystem of the mind/brain devoted to language acquisition via experience
(c) How is knowledge of language put to use?	Linguistic performance, the uses of language via experience.

[198] Churchland, *Braintrust*, 14.
[199] M. Tisak and E. Turiel, 'Children's Conceptions of Moral and Prudential Rules', *Child Development* 55 (1984): 1030–1039.
[200] Mikhail, *Elements of Moral Cognition*, 15.

Morality	
Question	Answer
(a) What constitutes moral knowledge?	Mind/brain has system of moral knowledge 'Sense of justice' 'moral sense', 'conscience', 'moral faculty' etc.
(b) How is knowledge of language acquired?	(UMG) Universal Moral Grammar: A theory of the initial state of the moral faculty, a subsystem of the brain, accounts for interaction via experience to reach moral maturity.
(c) How is knowledge of language put to use?	Moral Performance: How moral knowledge enters into actual representation an evaluation of human acts and other forms of behaviour.

Now if it is the case that children inherit a moral faculty in the brain's unfolding(s) as it is derived from its evolutionary history, then it may be found, even if in proto form only, in our nearest ancestral relatives. Indeed, it may even be more innate, in the sense of older and prior, to that of the language faculty itself.

Primatologist Frans De Waal's early work with chimpanzees indicates that 'the cognitive and emotional machinery of norm creation and norm enforcement was available long before language existed'.[201] Similarly, psychologist Johnathan Haidt's social intuitionist model argues that language arises so as to facilitate the depth of social engagement and culture**.** This is an intriguing and deeply interesting idea, but for now my concern is with innateness.

Hauser's interpretation of innateness refers to a *strict* evolutionary explanation:

> I suggest that what people generally mean when they debate the 'innateness of morality' is whether morality (under some specification) can be given an adaptive explanation in genetic terms: whether the present-day existence of the trait is to be explained by reference to genotype having granted ancestors reproductive advantage.[202]

One may talk about innateness in terms of a general learning capacity, or in terms of specific innate modules:

[201] Jonathan Haidt, 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment', *Psychological Review* 108, no. 4 (2001): 826.
[202] Richard Joyce, *Evolution of Morality*, 3.

> Sometimes the specification rests on the idea that for anything we can easily acquire through learning, the genes provide the brain with an innate capacity – structural readiness… On a more restricted use of innate it refers to those behaviours that are both genetically programed and universally displayed by all individuals that carry the relevant genes.[203]

Either way, as Chandra Sripada points out, two forms of nativism become possible:

- *Capacity nativism*, which refers to both theory of action (cognitive functionality, which allows one to distinguish between objects, break them down into parts, understand their relations causally etc.) and theory of mind (aligning conscious understandings with specific evolved functions such as cheater detection). An example of a theory of action is seen in Kleffner and Ramachandran's argument that there are inherited biological perceptual mechanisms for inferring the extension of objects in three-dimensional space.[204]
- *Content nativism,* which refers to whether there is an innate structure to the moral norms we have, so really what we are discussing in terms of content are certain actions automatically prohibited by an innate module.

Most authors have assumed capacity nativism implies content nativism.

Hauser's position has been summarised as an application of Chomsky's principles-and-parameters position.[205] Chomsky argued for an underlying universal syntactical grammar that can be modified within parameters – the head of a phrase, for instance, may have a parameter/switch which for some languages is on, others off, so that in one the head comes first in the phrase, in others at the end. The principles and parameters model adopted by Hauser adopts this model to account for innate structure in moral norms (e.g., 'no incest!'), while affording flexibility in how they are set (who counts as kin, and so forth).

[203] Churchland, *Braintrust*, 106.
[204] D. A. Kleffner and V.S. Ramachandran, 'On the Perception of Shape from Shading', *Perception and Psychophysics* 5,2 no. 1 (1992): 18–36.
[205] Scott M. James, *Introduction*, 108.

The underlying justification for such Chomsky/Hauser approaches is the 'poverty of stimulus' phenomenon. For it seems clear that when we acquire a language and/or a moral code, we gain something whose competent exercise far transcends what might be derived from the stimulus/inputs on which this acquisition rests. This seems possible only on the assumption of innate structure**.**

If applied to morality the principles/parameters approach allows for the difference in moral positions found between, and in, local cultures at the same time as explaining their universal and basic commonality. Richard Joyce takes this approach:

> The hypothesis which I argued did not deny that cultural learning plays a central role in determining the content of the moral judgments that an individual ends up making; the claim was just that there is a specialized innate mechanism (or series of mechanisms) whose function is to enable the type of acquisition. This mechanism, I hypothesize, comes prepared to categorize the world in morally normative terms; moral concepts may be innate if moral beliefs are not.[206]

This version of innate morality has received much criticism, though it is not unreasonable. Prinz, for example, argues that behavioural accounts of animal altruism (as, for instance, in de Waal), are simply not decisive:

> Morality, like all human capacities, depends on having particular biological predispositions, but none of these, I submit, deserves to be called a moral faculty. Morality is a by-product—accidental or invented—of faculties that evolved for other purposes. As such, morality is considerably more variable than the nativism program might lead us to think, and also more versatile.[207]
>
> … Bees are altruistic (they die to defend their colonies), but we do not say they have a moral sense. Innate good behaviour can be used as evidence for an innate moral sense, but… the evidence is far from decisive.[208]

Evolutionary psychologist Scott James makes a similar point:

---

[206] Richard Joyce, *Evolution of Morality*, 181.
[207] Ibid., 367.
[208] Ibid., 370.

> Our moral sense was, if you will, a 'by-product' of some other system that was directly selected for. As a point of comparison, consider the colour of human blood. No one seriously believes that the redness of human blood was directly selected for. What was directly selected for was the oxygen—carrying properties of blood; the redness 'came along for free'. That was an accidental property of blood. In the same way, some wish to claim that our moral sense was an accidental property of other cognitive adaptations—for example, our capacity to reason about the consequences of our actions…The point is that we should be careful not to conclude that a piece of behaviour is (or, more carefully put, is produced by) a psychological adaptation just because it happens to be biologically adaptive.[209]

Although we do not share a recent evolutionary history with bees, Prinz's claim is clear: in a deep sense 'morality is artificial all the way down'. Arguing against Mikhail and others who stress the poverty of stimulus idea, Prinz insists that children may indeed acquire their basic moral inclinations and distinctions in their first three years. Sripada expands on this, arguing that moral norms need not be an inductive learning problem, as children using language can simply be instructed into morality.[210] Thus while Sripada accepts linguistic nativism, he rejects moral nativism.[211] While there may be some bias towards an innate morality (e.g. in that we are less likely to murder for no reason), morality is largely determined by the environment the child lives in, and thus the experiences they have.

Sripada rejects moral norms conceived as 'a rule or principle that specifies actions which are required, permissible, or forbidden independent of any legal or social institution'.[212] On the grounds that none have been shown to exist:[213]

> There is solid ethnographic evidence that in domains such as social exchange (Fiske, 1991; Henrich etal, 2001), harms and violence (Robarchek & Robarchek,

[209] Scott M. James, *Introduction*.
[210] Chandra S. Sripada, 'Nativism and Moral Psychology: The Models of the Innate Structure that Shapes the Contents of Moral Norms', in Sinnott-Armstrong, *Moral Psychology*, 1:328.
[211] C. Sripada and S. Stich, 'A Framework for the Psychology of Norms', in *The Innate Mind,* Vol. 2, *Culture and Cognition*, ed. P. Carruthers, S. Laurence, and S. Stich (New York: Oxford University Press, 2006).
[212] Sripada, 'Nativism and Moral Psychology', 320.
[213] Ibid., 323.

> 1992; Keeley, 1996), hierarchy and social stratification (Bohem, 1999), marriage (Durham, 1991), sexual rules and prohibitions (Bourguinon & Greenbaum, 1973), food taboos (Simons, 1994) and many others, moral norms differ substantially across human groups.[214]

Given this substantial variability Sripada denies that moral norms can be explained in terms of the operation of principles that apply to language acquisition as such. He is prepared to admit there are certain *themes*, but denies these can be further specified. Instead he favours what he calls the 'innate bias model' to explain the innateness of moral norm content. The flexibility of this model makes plenty of room for cultural changes in moral norm content over time, without entirely repudiating nativism. It is a matter of probability claim over a population and probabilistic claims about an agent's decisions.

When it comes to the neuroscientific evidence for innate modules, Prinz is critical of the Jorge Moll,[215] who attempted the identification of moral circuits in the brain by comparing neuronal responses to pictures of moral scenes and non-moral scenes:

> First of all, the brain structures in question are implicated in many social cognition tasks, so we do not have reason to think they are specialized for moral appraisals. Second, there is considerable overlap between the moral picture condition and the unpleasant picture condition. Both cause increased activation limbic areas like the amygdala and insular cortex, as well as visual areas (due presumably to increased attention to the photographs).[216]

Churchland responds to this difficulty with modularity with a plural approach involving multiple areas of the brain, whose functions are not specific to morality. Her viewpoint rests on the inability to distinguish morality (hence the importance of defining morality in the first place) in terms of the brain's functionality:

> The classic problem that bedevils all innateness theories of behaviour is that in the absence of supporting evidence concerning genes and their relation to brain

[214] Ibid., 324.
[215] Jorge Moll, 'The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions', *Journal of Neuroscience* 22, no. 7 (2002): 2730–2736.
[216] Prinz, 'Is Morality Innate?', 386.

> circuitry involved, the theories totter over when pushed. Haidt for example, relies quite heavily on whether or not a skill is easily learned to demarcate skills the brain is innately 'prepared' for, and those that it is not so prepared for. But how do you defend, without resorting to ad hoc fixes, the innateness of some 'easily learned' things while excluding other 'easily learned' things, like riding a bicycle, tying a reef knot, putting on shoes?[217]

The upshot is that claims of the innate status of foundational moral norms are not conclusive, and we can appreciate why many prefer to talk loosely here of proportions, as in 'perhaps 70%, innate, 30% cultural', however vague and ambiguous they may sometimes be.

As my ultimate concern is with what a naturalistic and non-eliminative reductionist account of the moral mind can do for us when we approach environmental decision making and policy construction my concern with the question of innateness is finally a practical one. As Valerie Tiberius in a critical account of Prinz's work writes, at a moment like this we must 'marshal all of our psychological resources, whether or not they count as uniquely moral'.[218]

4.5 Emotion *v.* Reason

> *In the last decade, neuroscientists and psychologists have produced a substantial body of empirical evidence that challenges established views of morality and rationality. This evidence may be incompatible with the central methodology in practical ethics, which involves putting weight on intuitions in ethical reflection (Rawls 1951, 1972; Daniels 1996). Employing neuroimaging and psychological experiments, Haidt (2001), Hauser (2006) and others have documented unconscious influences on moral judgment with little input from consciousness.*[219]

The emotion *v.* reason dispute as to the ultimate origin(s) of morality is important to this thesis because if morality is strongly emotional in origin then we may be able to explain

[217] Churchland, *Braintrust*, 116.
[218] Valerie Tiberius, 'The Nativism Debate and Moral Philosophy: Comments on Prinz', in Sinnott-Armstrong, *Moral Psychology*, 1:426.
[219] J. Savulescu et al., 'Neuroscience of Morality', *Oxford Centre for Neuroethics*, accessed 1 March 2014, http://www.neuroethics.ox.ac.uk/research/area_4.

why, in the face of the Anthropocene challenge, we have largely as yet failed to respond adequately to what we know, rationally and on the basis of best reasoning, to be perhaps the greatest challenge of all.

If many in philosophy, influenced by Kant, have sought the roots of morality in reason, or pure practical reason, alone, many in the sciences, and particularly over the last couple of decades, have given a philosophical nod to Hume and looked to the emotions. Both approaches were explored in the twentieth century by American developmental psychologists: Martin Hoffman stressed the role of empathy in moral development, while Lawrence Kohlberg's developed the role of reason.

It is an ancient and entrenched prejudice that humanity is distinguished by its capacity for reason(ing). Given this exceptionalist conviction, if the origin(s) of morality can be found in primates, our evolutionary relatives, and it correlates to our own, this might suggest that morality is founded not in reason, but in the emotions. If then, as Warneken and Tomasello[220] discovered, children appear to display altruistic or even moral behaviours without being taught them, it plausibly follows they have inherited them, especially given they display these behaviours prior to learning language. As what it is they display is not the product of distinctive (human) reason or reasoning, and that it has definite motivational potency, it is all the more likely that our morality is a product of evolved emotional responses, capacities and incapacities.

De Waal supports this line of approach, arguing that seeking to tie the origins of morality to reason has the unfortunate implication that apparently altruistic behaviour is viewed as a veneer—an unnatural layer around a selfish core—with the further, even more unfortunate implication, that humans can only become moral through defiance (and denial) of our inner nature.

> According to veneer theory, we are ruthlessly self-interested creatures, who conform to moral norms only to avoid punishment or disapproval, only when others are watching us, or only when our commitment to these norms is not tested by strong temptation… morality then enters the scene as a set of rules that

[220] F. Warneken and M. Tomasello, 'Helping and Cooperation at 14 Months of Age', *Infancy* 11 (2007): 271–294.

> constrain this maximizing activity (of veneer theory)… veneer theory holds that these constraints, which oppose our natural and rational tendency to pursue what is best for ourselves, and which are therefore unnatural, are all too easily broken through.[221]

This opinion – that humans as moral beings are required to (somehow) break with nature – is prominent in evolutionary biology. Here is George C. Williams:

> I account for morality as an accidental capability produced, in its boundless stupidity, by a biological process that is normally opposed to the expression of such a capability.[222]

Philosophically, this kind of idea can be found in Freud and Nietzsche.[223]

De Waal launches a scathing attack on Veneer theorists such as Dawkins and Huxley. His basic point derives from Hume. For if reason does (and clearly it does) play a part in morality, then it does so only in the content of, and from the basis of, intuitional values. Reason, after all, must be *motivated*. That later evolutionary products, like our PFC, facilitate reasoning, is not in doubt. But what drives us to reason and along the lines we do, isn't itself a product of reason.

> David Hume argued moral judgements are derived similar in form to aesthetic judgements: They are derived from sentiment, not reason, and we attain moral knowledge by 'an immediate feeling and diner internal sense', not by a 'chain of argument and induction'…. Hume argued 'reason is and ought only to be the slave of the passions' (Hume 1777/1960, 2)[224] ….(Though by *passion* Hume meant something more general than emotion; he had in mind any practical orientation toward performing an action in the social or physical world.)[225]

---

[221] C. Korsgaard, 'Morality and the Distinctiveness of Human Action', in de Waal, *Primates & Philosophers*, 99.
[222] George C. Williams, 'Reply to Comments on "Huxley's Evolution and Ethics in Sociobiological Perspective"', *Zygon* 23 (1988): 438.
[223] Korsgaard, 'Morality', 104.
[224] Haidt, *Emotional Dog*, 814.
[225] Churchland, *Braintrust*, 5.

De Waal's work and empirical studies have breathed new life into Hume's insight, and done so in the context of stressing the continuities and parallels in human and primate behaviour. For de Waal it is clear that emotions are at the origin of human morality**,** and so are ontologically prior to that reasoning and language capacity that developed subsequently in intensely social environments.

His stress on the centrality of emotions is shared by Damasio:

> Emotions occupy a central role; it is well known that, rather than being the antithesis of rationality, emotions aid human reasoning. People can reason and deliberate as much as they want, but, as neuroscientists have found, if there are no emotions attached to the various options in from of them, they will never reach a decision or conviction (1994).[226]

Damasio is responsible for the neuroscientific somatic marker hypothesis (SMH). According to SMH, 'experiences in the world trigger emotional experiences that involve bodily changes and feelings'.[227] For Damasio emotions are changes in both body and brain states in response to stimuli. The theory was tested successfully by Batson, Engel, and Fridell (1999) and Wheatly and Haidt (2001). De Waal's argument is that this capacity, shared with primates, is the origin of morality. He sums his position up in three points:

> A. psychology: 'human morality has an emotional and intuitive foundation' B. Neuroscience: 'moral dilemmas activate emotionally involved brain areas'. C. Primate Behaviour: 'Our relatives show many of the tendencies incorporated into human morality'.[228]

While de Waal's argument is compelling, it seems clearly to stand in some tension with Darwin's own views. For Darwin 'a moral being is one who is capable of comparing his past and future actions or motives, and of approving or disapproving them. We have no reason to suppose that any of the lower animals have this capacity'.[229]

---

[226] de Waal, *Primates & Philosophers*, 18.
[227] Haidt, *Emotional Dog*, 822.
[228] de Waal, *Primates & Philosophers*, 22.
[229] Charles Darwin, *The Descent of Man, and Selections in Relation to Sex* (1871), 88–9.

The underlying point is well captured by the modern Kantian, Christine Korsgaard. Emotions – affective states – are, she feels, insufficient for morality and so, moral judgement, deliberation and action, for such things presuppose rational choice (autonomy):

> The animal's purposes are given to him by his affective states: his emotions and his instinctual or learned desires…even where [the animal] does choose the choice is made for [the animal] by the strength of his affective states.[230]

For Korsgaard what makes human beings different to animals is that we are capable of 'caring about what we are, not just what we can get'.[231] In this resolutely Kantian interpretation, morality is grounded in a style of reasoning (one bound by some idea or notion, however limited, of universalizability and impartiality), and so is unavailable to non-human animals.

I reject this straight-forward 'refutation'. It ignores, or underplays, the essential motivating role of the emotions. It does the same with the cognitive dimension of emotions (fear is of that which is really fearful, otherwise it misfires). It gives us a conception of morality that demands universalizability and impartiality, overlooking that for these demands to generate morality (and not, say, universal indifference or the frightening consistency of the sociopath) they must operate from within an empathic perspective. And it is simply wrong to think of all, let alone most, of moral life in terms of choice or decision. As Simone Weil pointed out, the man who plays with his son because 'it is my duty', and he has decided he must fulfil his duty, is not at all a better man morally than the man who plays with his son 'because I love playing with him.'

Like de Waal (though not as willing as he to fully attribute morality to the primates) Haidt argues for a kind of moral cognition that is distinct from reasoning, though one in which reasoning has a place, if not at all the place Korsgaard and the Kantians think.[232] For Haidt pretty much all distinctively moral reasoning (as opposed to calculative reasoning) occurs after the fact, as rationalisation not primal causation.

---

[230] Ibid., 110.
[231] Ibid., 119.
[232] Haidt, *Emotional Dog*, 814.

I think it reasonable to conclude that emotions underlie much of our decision making, whether directly or via a level of emotional reasoning distinct from conscious moral reasoning. While this position is at least as old as Hume, it has not often been applied to issues of environmental ethics.

There has been some work. Basic survey testing was performed by Carmen Tanner in 2009. Tanner uses the term 'sacred values' to refer what we earlier called intrinsic value. She found that rather than utilitarian reasoning driving deliberation and decision, sacred values appeared to do most of the work.[233] Such sacred or intrinsic valuation seems obviously a matter more of emotional connection or reaction than anything simply cognitive.

To develop further my naturalistic non-eliminativist reductionist approach, I turn now, as I must, to the field of evolutionary psychology.

[233] Carmen Tanner, 'To Act or Not to Act: Nonconsequentialism in Environmental Decision Making', *Ethics and Behaviour* 19, no. 6 (2009): 491.

Chapter 5: Evolutionary Psychology and Heuristics

5.1 Introduction

> *Recently, researchers have begun to look for moral modules in the brain, and they have been increasingly tempted to speculate about the moral acquisition device, and innate faculty for norm acquisition akin to celebrated language acquisition device, promulgated by Chomsky.*[234]

We have discussed the view that our brain is composed of various modules each serving a specific function. We have seen this view is not accepted by everyone (e.g. Churchland, with her pluralist account), but it is certainly the dominant approach when it comes to evolutionary psychology. The strength of the idea lies in the plausibility of a piecemeal approach to the brain's construction. It seems unlikely that humanity evolved or mutated the full plasticity of mind we now enjoy, over a relatively short period of time. For psychologists the challenge is to explain this in terms of the operations of natural selection over geological time.[235] Evolutionary theory as applied to features of animals (legs**,** wings etc.), is reasonably simple to comprehend, but can the same be said for the composition of our brain and the capacities it expresses?

> You may have no problem accepting a Darwinian explanation for the structure of the human eye. Ditto for the human lungs, liver, colon, and circulation system. But what about jealousy? What about friendship? What about men's proneness to violence, or women's interest in looking young? What about language? These things, you say, are another matter. Perhaps not, say evolutionary psychologists.[236]

For evolutionary psychologists an appropriate analogy to explain the human mind is that of the body. Evolutionary psychology is thus distinct from cognitive psychology.[237] It is

[234] Prinz, 'Is Morality Innate?', 367.
[235] Viren, *Evolutionary Psychology*, 32.
[236] Scott M. James, *Introduction*, 18.
[237] Stephen M. Downes, 'Evolutionary Psychology', in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford University, Fall 2010), http://plato.stanford.edu/archives/fall2010/entries/evolutionary-psychology/.

foremost a theory of cognitive architecture,[238] not an account of our cognitive experience. Evolutionary psychology attempts to reconstruct the mind's design from an analysis of the problems the mind must have evolved to solve.[239] This includes an account of a moral mind.

Leda Cosmides and John Tooby began developing such an account in the early 1980s with their 'social contract theory'.[240] Their central concern was with the type of reasoning that develops through situations of social exchange. The underlying methodological commitments may be expressed this way: we apply evolutionary theory (adaptationism) to localised elements of the brains architecture (modules), which we associate with certain behavioural modalities. We then reverse-engineer our way to an evolutionary 'just so' story of that development, theorising the type of scenarios that may have led to specific localised modules. In doing this we enrich and complexify the causal story. From claims/facts about gene traits, and more generalised theories of cooperation and punishment from evolutionary circumstances, we can now move on to understanding specific functional classes that are crucial to morality, like cheater detection.

If we are to apply evolutionary theory to understand our psychology, then, as David Wilson points out, there are four rules we should heed:[241]

- Proximate and ultimate causation need to be studied in conjunction with each other.
- Proximate mechanisms need not resemble functional consequences in anyway whatsoever. They merely need to cause the functional consequences.
- Cultural evolution plays a strong role in the evolution of proximate mechanisms that motivate altruistic actions, not just genetic evolution.
- The evolutionary fate of a given psychological mechanism that leads to altruistic action depends critically on the environment, including the human constructed environment.

[238] Ibid.
[239] D. Buller, *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature* (Cambridge, MA: MIT Press, 2005), 92.
[240] Cosmides, 1985.
[241] David Sloan Wilson, *Does Altruism Exist?*, 68–70.

5.2 Modules and Logics

There is no single definition in evolutionary psychology of 'module', and some have even argued for its dismissal as it implies compartmentalisation of the brain itself.[242] (As we will see, this challenge is based on a misrepresentation.) What we find, in the literature, are four non-exclusive ways of thinking about modules. There is the *Neural Module* (the relation between neural networks and a bodily action); the *Perceptual Module* (the coordinator of neural modules when in interaction with the environment); the *Cognitive Module* (our ability to conceptualise information and relations utilising information from other modules theoretically and imaginatively); and the *Functional Module* (systems which integrate other modules to determine behaviours).[243] Behind all of these conceptualisations there is a common idea, captured by David Buss, when he defines a module as 'a set of procedures within the organism that is designed to take in a particular slice of information and transform that information via decision rules into output that historically has helped with the solution to an adaptive problem'.[244] I adopt this simple view for the purposes of this enquiry. It follows from this specification that we may well have a module for (say) walking, but not for buying handbags.

Tooby and Cosmides are pioneers in the development of evolutionary psychology. For them the brain is a biologically designed and realized computer. They summarize their view as follows:[245]

- The brain is a computer designed by natural selection to extract information from the environment.
- Individual human behaviour is generated by this evolved computer in response to information it extracts from the environment. Understanding behaviour requires articulating the cognitive programs that generate the behaviour.

[242] Viren, *Evolutionary Psychology*, 34.
[243] D. C. Geary and K. J. Huffman, 'Brain and Cognitive Evolution: Forms of Modularity and Functions of Mind', *Psychological Bulletin* 128 (2002): 698.
[244] David Buss, *Evolutionary Psychology and the New Science of the Mind* (Boston: Allyn and Bacon, 2007), 52.
[245] Tooby and Cosmides, 'Conceptual Foundations', 18.

- The cognitive programs of the human brain are adaptations**.** They exist because they produced behaviour in our ancestors that enabled them to survive and reproduce.
- The cognitive programs of the human brain may not be adaptive now; they were adaptive in ancestral environments**.**
- Natural selection ensures that the brain is composed of many different special purpose programs and not a domain general architecture.
- Describing the evolved computational architecture of our brains allows a systematic understanding of cultural and social phenomena.[246]

The fourth point (that the cognitive programs of the human brain may not be adaptive now, though they were originally) is of particular interest as it marks a split between the views of evolutionary psychologists and cultural evolutionists. The Santa Barbara School of evolutionary psychology holds to an 'unchanging set of cognitive adaptations',[247] while cultural evolutionists (often drawing on epigenetics) hold that such adaptations are subject to the pressures of natural selection embodied in culture itself.

The argument for massive modularity is straightforward:

> 1. The human mind is a product of natural selection. 2. In order to survive and reproduce, our human ancestors had to solve a range of adaptive problems (finding food, shelter, mates, etc**.)** 3.Since adaptive problems are solved more quickly, efficiently, and reliably by modular (domain-specific, mandatory, etc.) systems than by non-modular ones, natural selection would have favoured the evolution of a massively modular architecture. 4. So the human mind is probably massively modular.[248]

In 2012 support for the massive modularity thesis arose on the grounds of the lower short-term energy expenditure required to develop modular networks through time:

[246] This theory has its roots in Jerry Fodor's 'massive modularity thesis'.
[247] Tim Lewens, '*Cultural Evolution: Integration and Skepticism*', in *The Oxford Handbook of Philosophy and Social Science*, ed. H. Kingcaid (New York: Oxford University Press, 2012), 2–17.
[248] Philip Robbins, 'Modularity of Mind', in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford University, Summer 2010), http://plato.stanford.edu/archives/sum2010/entries/modularity-mind/.

> The ubiquitous direct selection pressure to reduce the cost of connections between network nodes causes the emergence of modular networks.[249]

What are our brains moral modules? One instance, often pointed to, lies in our natural behavioural inclination to protect our relatives. However, the most researched module is Cosmides and Tooby's Cheater-Detection-Module, [250] born from such reasoning experiments as the Wason Task,[251] later expanded to account for other aspects of our moral behaviour.[252] The adaptive problem to which this module provides a solution is detecting cheaters in situations where there is a social exchange (of resources). Biologists claim that if we could not detect cheaters, we could not evolve cooperative traits (because of what Dawkin's refered to as 'subversion from within').[253] Tooby and Cosmidies have taken this further, suggesting cheater detection module potentially supports an innate moral capacity as it suggests the existence of a specialized modular reasoning system.[254]

Further support for modularity (though here in the network sense as opposed to cognitive sense) has arisen in the form of mathematical models showing that cooperation can evolve and remain stable if individuals punish both cheaters and those who fail to punish cheaters.[255] In a like manner, mathematical models focused on the prisoner's dilemma have done the same thing.

So how might such a module arise as the product of natural selection? We would need variation – say of moral decisions/choices in defined situations (for instance, PD situations) – the heritabilility of that which grounds these variations, and differential reproductive rates resulting from such inherited variation. With this in place, and over geological time scales, we can imagine the architecture of the brain as we now know (and have) it arising.

---

[249] Clune, Lipson, and Mouret, 'The Evolutionary Origins of Modularity', *Proceedings B: Biological Sciences*, last modified 13 January 2013, http://rspb.royalsocietypublishing.org/content/280/1755/20122863.
[250] Leda Cosmides and John Tooby, 'Cognitive Adaptations for Social Exchange', in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, ed. J. H. Barkow, L. Cosmides, and J. Tooby (Oxford: Oxford University Press, 1992).
[251] P. C. Wason, 'Reasoning', in *New Horizons in Psychology: I*, ed. B. M. Foss (Harmondsworth, England: Penguin, 1966).
[252] Cosmides and Tooby, 'Cognitive Adaptations'.
[253] J. Van Lier, R. Revlin, and W. De Neys, 'Detecting Cheaters without Thinking: Testing the Automaticity of the Cheater Detection Module', *PLoS ONE* 8, no. 1 (2010): e53827, doi:10.1371/journal.pone.0053827.
[254] Prinz, 'Is Morality Innate?', 384.
[255] M. D. Hauser, *Moral Minds*, 432, quoting Boyd and Richerson, *Not by Genes Alone*.

It has been argued that the stimuli that trigger such modules in social exchange do not need to occur in real life; they can occur in imagination[256] or meta-representations.[257] (This means that deliberative reasoning is not immune from the influence of domain-specific evolved programs).[258]

If this is on the right track then we can see how morality (moral choice, decision and deliberation) might be explained by evolutionary psychology.

We can look at the architecture proposed by the evolutionary psychologist, in terms of what these modules produce. In the case of morality, that is a brain compartmentalised by deontic logic rules. As deontic reasoning seems to split up into functionally distinct domains (different modules), and because of the way in which natural selection tends to engineer evolved systems,[259] it may not be possible to pursue a domain-general deontic logic, as there may be no uniform operators across all contexts involving human interaction.[260]

Much of the experimental testing undertaken in support of evolutionary psychology has focused on identifying deontic domains. In particular, evolutionary psychologists have noted that reasoning takes the form of two systems (though there are likely more); one for *social exchanges*, the other for *precautionary rules*. The former contains a myriad of social contract algorithms that may be triggered by the specific use of certain words, or just actions.[261] Neuroscience has correlated with this theory, showing different brain areas are utilised for social exchange reasoning than are used for precautionary rule reasoning. (Precautionary rules present the template: *If one is to engage in hazardous activity, then one must take precaution*.) Given this we can begin to understand the 'cooperative story' a little better, mapping out and moral exchange in terms of deontic logics ('If you give me Q, I will give you P'), and social contracts ('We both benefit if we trust each other' etc.) We

[256] P. Boyer, *Religion Explained: The Evolutionary Origins of Religious Thought* (New York: Basic Books, 2001).
[257] A. M. Leslie, 'Pretense and Representation: The Origins of "Theory of Mind"', *Psychological Review*, 94 (1987): 412–426.
[258] Cosmides and Tooby, 'General Deontic Logic', 56.
[259] Gallistel, 2000.
[260] Cosmides and Tooby, 'General Deontic Logic', 57.
[261] Ibid., 113.

can also understand the evolution of mutually beneficial behaviours, like 'cheater detection':

> By seeing what conditions turn the cheater detection mechanism on and off we found that its procedures define a cheater as an individual who has (i) taken the benefit the provider agreed to supply contingent on a requirement being satisfied, (ii) done so without having satisfied the providers requirement, (iii) taken these actions by intention rather than accident.[262]

Expansions on this style of account utilise neural functionality. Peter Ulric Tse provides an extended account of how 'cross-module binding' might explain our capacity for analogical reasoning (a common kind of moral reasoning). Cross-module binding would explain why we are reminded of something when we are viewing something else; or, to put it more precisely, it would explain semantic representations through multiple nodes in a network of activated neurons.

> Analogical thought is a by-product of the synchronous firing, coupling, or entrainment of distal neural populations across modules…The emergence of binding across modules triggered not only the birth of symbolic thought by permitting arbitrary objects to stand for arbitrary referents but also triggered automatic cross-modular activations.[263]

At the heart of Tse's argument is the idea that through cross modular binding humans developed the ability to form associations across previously encapsulated classes of information.[264] And this ability, Tse argues, preceded language:

> The ability to use and recognize symbols must have preceded the evolution of language because whereas syntactic processing is necessarily symbolic, symbolic processing need not be syntactic. Language by its very nature involves the utterance and manipulation of symbols.[265]

---

[262] Ibid., 114.
[263] Peter Ulric Tse, 'Symbolic Thought and the Evolution of Human Morality', in Sinnott-Armstrong, *Moral Psychology*, 1:281.
[264] Ibid., 276.
[265] Ibid.

Tse argues symbolic thought gave rise to the possibility of morality, and so arguing (for instance) that a monkey lacks morality because it lacks symbolic cognition:

> Once acts became symbolized, they could now stand for, and be instances of, abstract classes of action such as good, evil, right, or wrong. Symbolic thought permitted new dimensions of behavior, for example, the expression of territoriality over the ownership of an idea rather than just a concrete thing such as turf, a bone, or a mate. Thus, while a monkey has affection, social intelligence, likes, dislikes, fear, inhibitions, territoriality, deceit, aggression, vengefulness, and other predispositions that govern behavior, these are not morality.[266]

For Tse an act is judged immoral because 1) that act in the brain is represented by a symbol, 2) that symbol is a member of an abstract category, and 3) that category is one of disvalue. So an act, for example kissing a friend's partner, becomes more than the act of kissing, rather act is conceived under the abstract category of 'wrong'.

How might an account like Tse's impact on Anthropocene concern that we develop attitudes towards natural entities and processes that facilitate sustainability? Well, do we have any reason to think we possess an automatic emotional prohibition defined by a specific module that prevents us from harming the environment? Certainly this is what the Routley's suggest with their "Last Man Argument"[267], contending that pretty much all of us think that it would be wrong for the Last Man to destroy the natural world; and it is certainly true that many of us experience feelings of remorse or sadness upon witnessing environmental catastrophe, just as some of us (myself, for instance) find viewing an axe cutting into a tree and the sap oozing out sickening, even disgusting. How might this have come about? Scott James outlines a plausible answer when he discusses the evolution of morality:

> If a solution to an adaptive problem can be had on the cheap, chances are, natural selection will take it… jerry rigging new solutions out of old structures, we may well have inclusive fitness to thank for putting core psychological system in place that made later moral (or quasi moral) behaviour possible… Since early humans were already disposed to care about those closest to them (thanks to inclusive

[266] Ibid.

[267] Routley, Richard. 1973. Is There a Need for a New, an Environmental, Ethic? Proceedings of the XVth World Congress of Philosophy 1: 205-210

> fitness), it's not so difficult to imagine a few more mutations, aided by regulation environmental pressures, delivering a disposition to care about a much wider range of folk *and fauna and flora*.[268]

Whatever the final truth here, the general story is reasonable: environmental ethics might be an extension of existing morality (we have the cultural software in pre-existing modular hardware). It is not (or may not be) that we have particular module for environmental welfare, (as, perhaps, with cheater detection), but rather existing capacities are activated by informational stimuli. Such information would have to come in a certain form (an appropriate cue) to be utilised. It follows that situations which present different stimuli or cues may be picked up by various existing modules. Given this, it follows that if we want to understand behaviours of populations we need minimally to understand the evolved brain's modules—and evolutionary psychology gives us this potential.

5.3 Issues

> *… evolutionary psychological accounts of human behaviour are like polls in the sense: they measure large-scale trends. They predict what most humans will be like. Actually, such accounts are more general than even this. Evolutionary Psychological accounts predict what most humans will be like under specific circumstances [only].*[269]

There are certain reductionist complaints with this modularity approach, typically from neuroscientists.[270] Uttal, for instance, regards the modularity thesis as probably false. He lists five reasons:[271]

- The neutrality of formal models;
- The neutrality of behavioural data and findings;
- The inaccessibility of mental activity;
- The un-analysability of mental activity; and
- The complexity of both mental activity and its relative neural underpinnings.

[268] Scott M. James, *Introduction*, 39.
[269] Ibid., 24.
[270] J. Dupré, *The Disorder of Things: Metaphysical Foundations of the Disunity of Science* (Cambridge, MA: Harvard University Press, 1993).
[271] Uttal, *New Phrenology*, 145.

On the basis of such points D. Buller (2005) has argued that it is the brain's 'developmental plasticity' that is responsible for seemingly the distinct functions of the mind, rather than historical evolutionary environments, and that the empirical evidence cited by evolutionary psychologists is flawed. Though Buller accepts the constructivist evolutionary history of the brain, he refuses the piecemeal approach characteristic of modularity theorists.

While one can appreciate the force (though not the finality) of such critical remarks, most in this area favour evolutionary psychology. As Pinker writes: 'Saying that the brain solves these problems because of its "plasticity" is not much better than saying it solves them by magic'.[272]

A further critical argument is that even if the modules of the brain the result of environment specific construction, still, because of the mass action nature of our brain functions, it may be impossible to reduce it this way to specific evolutionary events (and in that case, we may as well revert to cognitive psychology). Furthermore, there are few examples of evolutionary modules, as much as we might hypothesise and prove through behavioural tests the existence of those deontic logics they supposedly produce.

Still, I think that utilising work like Tse's, we can arrive at more than a 'just-so' story, or at least a really good one. My basic concern is with environmental stability, but as Wallace says with other ends in mind:

> If there is a kind of tokenisation that does operate with racist, sexist, and discriminatory behaviour, then the theory may be helpful in developing methods in moral education that aim to disrupt the processes which lead to such attitudes.[273]

5.4 Heuristics

[272] Pinker, *How the Mind Works*, 75.
[273] K. Wallace, 'Morality and the Capacity for Symbolic Cognition: Comment of Tse', in Sinnott-Armstrong, *Moral Psychology,* 1:310.

> *Human knowledge and its products, knowledge as commonly understood, are not unlike the knowledge gained through the intelligences of other species that have evolved secondary brain heuristics. All intelligence, including that of ourselves, is innately specified and domain-specific. We are no more possessors of a tabula rasa than are the birds and the bees.*[274]

The study of heuristics in psychology began with work by Amos Tversky and Daniel Kahenman in 1974. In philosophy Dewey referred to heuristics as 'habits', and other philosophers have often referred to them as 'intuitions'. The study of heuristics in psychology is yet another emerging field that seeks to explain our behaviours by representation of the reasoning processes, conscious or unconscious that our brain performs. To focus on heuristics is, when it comes to modularity, to focus on the 'product side', with the obvious assumption that such heuristics have an evolutionary foundation[275].

Say that we do have a cheater detection module thanks to natural selection favouring a capacity enabling us to (often enough) detect those who seek to take advantage of our cooperative cognition. As Gerd Gigerenzer points out, this module involves the downstream 'intuitive search rule'; a moral heuristic that seeks to reveal if one has been cheated in a social contract.[276] What we find here is a crossing of the divide between evolutionary biological origin, in terms of the modules, and conceptual understanding in terms of the type of information they present to consciousness. (This fits nicely into the idea that choice is an illusion and UCD reigns per Section 2.3).

What is missing is a neuroscientific account to support the legitimacy of a modularity centred evolutionary psychology. Certainly the matching process seems empirically reasonable. We will explore this representational relation in terms of the underlying neural mechanics later (in the cognitive neuroscience section). For now it is worth considering these 'front end' heuristics. The promise of doing this is obvious: if our brains determine which environmental stimuli provoke which brain heuristics and how they are prioritised relative to each other, then we can use this knowledge when considering policy design.

---

[274] Ibid.
[275] E.g. E. Sober and D. S. Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behaviour* (Cambridge, MA: Harvard University Press, 1999).
[276] W. Gaissmaier and G. Gigerenzer, 'Heuristic Decision Making', *Annual Review of Psychology* 62 (2011): 478.

> The study of heuristics will never replace the need for moral deliberation and individual responsibility, but it can help us to understand which environments influence moral behaviour and how to possibly modify them to the better. One and the same heuristic can produce actions we might applaud and actions we condemn, depending on where and when a person relies on it… Knowing the heuristics that guide people's moral actions can be of help in designing change that might otherwise be out of reach… The descriptive goal of the heuristics program is to spell out what the heuristics underlying intuition are, and how they differ from the post hoc rationalization of one's judgment.[277]

One of the central difficulties in the study of heuristics is addressing the question of how heuristic parameter values are selected in each environment.[278] In answering this question scientists typically invoke: a) an innate parameter, b) learning theory, c) imitation and teaching/social/cultural processes, and d) the content of one's memory. The next move is to consider the selected heuristics success or failure in an environment (this component is referred to as a heuristics 'ecological rationality').[279] Finally, to avoid the non-reductionist and explanation diluting temptation of 'identifying' a heuristic for every 'behavioural output', we need to follow the lead of Goldstein and Gigerenzer, who identify four classes of heuristics to replicate actual decision-making processes.

> The first class is recognition-based heuristics utilising memory. The 'recognition heuristic' is the primary heuristic in this class. It states that, 'if one of two alternatives is recognised and the other is not, then infer that the recognised alternative has the higher value with respect to the criterion'.[280]

The higher the *recognition validity*, x, the more ecologically rational it is to rely on it, and so the more likely people will rely on it– thus x = C/(C+W) where C is the number of correct inferences the recognition heuristic would make, and W is the number of wrong inferences.[281] This heuristic has been tested in election prediction, sports prediction,

[277] Gigerenzer, 'Moral Intuition', 15.
[278] Gaissmaier and Gigerenzer, 'Heuristic Decision Making', 456.
[279] Ibid., 451.
[280] D. G. Goldstein and G. Gigerenzer, 'Models of Ecological Rationality: The Recognition Heuristic', *Psychological Review* 109 (2002): 75–90.
[281] J. G. Johnson and M. Raab, 'Take the First: Option Generation and Resulting Choices', *Organizational Behavior and Human Decision Processes* 91 (2003): 215–229.

investment prediction and consumer choice among others, and in each case been found more reliable than standard regression techniques. For example, Ortmann et al. showed that recognised stocks outperformed on average stock experts and certain managed funds.[282]

Another heuristic in this class is the *fluency* heuristic that follows on from the recognition heuristic. So, 'if both alternatives are recognised but one is recognised faster, then infer that this alternative has the higher value with respect to the criterion'.[283] This applies to perceiving two things and to recalling from memory.[284] A further class of heuristics is referred to as '*one reason heuristics*'. Within this class is found the *take-the-best* heuristic, which shows how people infer which of two alternatives has a higher value on a criterion, based on binary cue values received from memory. This process has three elements:

> 1. The search rule: Search through cues in order of their validity. 2. Stopping rule: Stop on finding the first cure that discriminates between the alternatives (i.e. cue values are 1 and 0). 3. Decision Rule: Infer that the alternative with the positive cue value (1) has the higher criterion value.[285]

Heuristics can be *social* or *non-social*, in much the same way as they can be *moral* or *immoral***.** An example of a straight social rule is the '*default heuristic*' ('If there is a default in a social contract, do nothing about it').[286] Such defaults are set by institutions'.[287] (A fact that emphasizes the practical importance of setting defaults for environmental sustainability through our institutions). This reinforces the importance of local culture when it comes to utilising our innate decision-making hardware, for social heuristics, at least in terms of the stimuli with which they operate, may be intentionally manipulable. If certain heuristics in certain environments gives rise to immoral behaviour, equally there is great promise in developing moral heuristics.[288]

---

[282] A. Ortmann et al., 'The Recognition Heuristic: A Fast and Frugal Way to Investment Choice?' in *The Handbook for Experimental Economics Results*, ed. C. R. Plott and V. L. Smith (Amsterdam: Elsevier, 2008), 1(28): 993–1003.
[283] Gaissmaier and Gigerenzer, 'Heuristic Decision Making', 460.
[284] Ibid., 462.
[285] Ibid., 464.
[286] Ibid., 472.
[287] Johnson and Raab, 'Take the First', 215–229.
[288] C. R. Sustain, 'Moral Heuristics', *Behaviour Brain Science* 28 (2005): 531.

> Utilitarians ought to be able to identify heuristics for the maximization of utility; deontologists should be able to point to heuristics for the proper discharge of moral responsibilities; and those uncommitted to any large-scale theory should be able to specify heuristics for their own more modest normative commitments.[289]

Evolutionary psychology and heuristics might offer useful information too on our economic decision making as these impact on the environment. Those modules that create the demand for continued growth, and the heuristics used in the pursuit of such, are not just going to disappear**.** This means we must learn how to satisfy them in environmentally sustainable ways. Given, as Norton says, that–

> … environmental problems are characterized by ideologically committed advocates on both sides the problems are experienced as zero-sum competitions. The two sides in this polarized rhetoric, confused by miscommunication, resort to enmity and name-calling, and very little gets done to protect the environment. Seldom do environmental agencies and departments pursue coherent goals; and almost nothing is done efficiently once goals are chosen.[290]

– it is clear that the promise of this kind of enquiry has great potential value.

[289] Ibid., 532.
[290] Norton, *Sustainability*, x.

Chapter 6: Incorporating Culture

6.1 Introduction

Environmental decisions are obviously conceptualised in terms of the parameters of risk deemed socially acceptable. This means inquiry here has to go beyond the descriptive accounts provided by formal risk analysis and look at *culture*.[291] We need to understand where possible, the biological interface for cultural change and influence. One option is to model ideas as if such ideas were physical traits, and so in terms of evolution though natural selection. This was the idea of William James.[292] Modern examples of such theorising include the Dawkins/Dennett theory of Memetics, as well as the field of cultural evolution. Of course, there are critics of this strategy, and from within the field of evolutionary studies itself. Stephen Jay Gould argues that evolution as theory should not be applied to our ideas, as ideas do not amount to evolutionary 'individuals'.[293] However, cultural evolutionist Alex Mesoudi argues that this criticism is now invalid,[294] as geneticists have found that horizontal gene transfer is common in both bacteria and plants, with genetic material being transmitted across species.

> We now know that chimpanzees, orang-utans and cetaceans transmit non genetic information across generations, including forms of greeting, grooming and foraging, a characteristic of human culture.[295]

Henry Plotkin helped popularise these ideas in the early 1990s with his attempted reduction of knowledge types to fit within a Darwinian framework. In 'Darwin's Machines' he argued that genes and ideas are both forms of *knowledge*, and as such, subject to evolutionary processes. Plotkin's 'primary heuristic' is genetic knowledge, and his

[291] Sahotra Sarkar, *Environmental Philosophy: From Theory to Practice* (John Wiley & Sons, 2012), 68.
[292] William James, *The Principles of Psychology* (New York, NY: Henry Holt, 1890).
[293] Gould, *Structure*, 720; Gould, *Wonderful Life: The Burgess Shale and the Nature of History* (New York, NY: W. W. Norton, 1990), 63–5.
[294] Mesoudi, *Cultural Evolution*, 86.
[295] Viren, *Evolutionary Psychology*, 56.

'secondary heuristic' is a constrained, short-term, knowledge gaining process.[296] Both, he argues, operates via evolutionary processes.[297] Plotkin makes the reductionist argument that adaptations are in essence forms of biological knowledge about the external world:

> Adaptations are knowledge… [and] there are two features of adaptations… (1) each and every adaptation is for something… (2) Every adaptation comprises organization of an organism relative to some feature of environmental order… but why take the further step of equating adaptations with knowledge? Why knowledge, of all things, when it has a well-accepted common sense meaning and such a central and sensitive place in philosophy? How can the wing markings of a moth be knowledge? ... When you say that Europe is in the northern hemisphere, that knowledge comprises two components: a brain state, which is part of a organismic organization, and the world itself, which is the feature of environmental order relative to which that brain state stands. All human knowledge has the same two component relationship that adaptations have…. The connection I am arguing for is that all adaptations are instances of knowledge, and human knowledge is a special kind of adaptation.[298]

Plotkin's point is that *capacity nativism* potentially influences *content nativism.* He reduces information and ideas to a level ascribed by evolution that can synchronise, or run side by side with, the recognised common level of genetic evolution. Such reduction to one naturalist frontier offers the promise of decoding those causal factors that arise from *cultural evolution.*

> Cultural evolution is not a thing of the past. It is operating today faster than ever before, and unless mechanisms of coordination and protection are in place, cultural evolution will lead to outcomes that are highly dysfunctional for society as a whole.[299]

6.2 The Theory of Memetics

[296] Plotkin, *Darwin Machines*, 163.
[297] Ibid., 167.
[298] Ibid., 116.
[299] David Sloan Wilson, *Does Altruism Exist?*, 109.

Memetics is one example of applying the biological process of evolution to the development and history of our ideas. Richard Dawkins uses metaphors for describing genetic evolution ('replicator', 'interactor') and applies them to memes/ideas.[300] Dawkins sees selection at this level not in terms of its ontological objective construct (i.e. evolutionary psychology), but at the level of the idea (meme) itself. For Dawkins memes have similar capacities as the gene, with replication being the primary factor.[301] This is how cumulative cultural change occurs. Thus he applies the principles of fidelity, fecundity and longevity found in genetic evolution to cultural evolution.

In criticising memetics, Tim Lewens highlights three key issues with this application:

> *Cultural units are not replicators*: Replicators, remember, are supposed to be units that make copies of themselves. Some critics of the meme concept argue that there is no known mechanism that could explain how memes are copied. Mirror neurons have perhaps provided one answer to this objection, but the evidence is not decisive. *Cultural units do not form lineages*: A closely-related criticism of memetics draws on the fact that while in genetic replication we can trace a new copy of a gene back to a single parent, ideas are rarely copied from a single source in a way that allows us to trace clear lineages (Boyd and Richerson 2000)… *Culture cannot be atomised into discrete units:* Ideas stand in logical relations to each other. [302]

As I think there is an alternative to memetics that avoids these issues by not investing so heavily in a gene's eye view of evolutionary theory, I will not linger here. The alternative is found in *cultural evolution*:[303]

> Rather than seeking to show that there are cultural replicators, one can instead seek to build models that allow for error-prone learning, and that acknowledge

[300] Dawkins, *Selfish Gene*.
[301] J. Henrich, R. Boyd, and P. J. Richerson, 'Five Misunderstandings about Cultural Evolution', *Human Nature* 19 (2008): 120.
[302] Tim Lewens, 'Cultural Evolution', in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford University, Spring 2013), http://plato.stanford.edu/archives/spr2013/entries/evolution-cultural/.
[303] Henrich, Boyd and Richerson, 'Five Misunderstandings', 124.

that an individual's beliefs are often the result of exposure to many sources, rather than copying from just one source.[304]

6.3 Cultural Evolution

6.3.1 Introduction

> *We have read reports about genes for violence, homosexuality, alcoholism and even one for language. The essentialist stance is strengthened by the fact we have a tendency to believe that we are our brains. However, longstanding studies of reorganisation after injury have amply demonstrated that any such reductionist view of complex phenotypes is incomplete without consideration of intervening external and cultural factors.*[305]

Gould was prepared to accept levels of evolution only in so far as they remained on the biological level. But Plotkin argues that evolutionary processes operate at the level of our culture impacting on, shaping and selecting, the realm of information and ideas, and so on our behaviour. As Saniot puts it:

> For E. O. Wilson (2006) the brain in the process of evolution has predisposed humans towards co-operative behaviour. In other words, a bias evolved favouring cultural evolution that expressed moral codes of behaviour. The mind was probably guided by a litany of moral sentiments to the land and its organism, which humans had evolved with which eventuated as religious systems.[306]

Unlike memeticists, proponents of cultural evolution see no requirement for a strict analogue to evolution at the gene level ('cultural variants need not be particulate').[307] This is not a strict reverse engineering theory as is memetics; rather it recognises the gap between biology and behaviour, and attempts to apply evolutionary theory to cultural transmission between brains:

> An important theme emerging from this work concerns the dynamics of cultural evolution; for example, that cultural evolution can happen much faster than

[304] Lewens, 'Cultural Evolution'.
[305] Illes and Racine, 'Imaging or Imagining?', 11.
[306] A. Saniots, 'Making Connectivities: Neuroanthropology and Ecological Ethics', *Neuroquantology* 8, no. 2 (2010): 203.
[307] Ibid., 88.

biological evolution, and that cultural intuitions can constitute a change in ecological conditions that in turn can alter selection pressures.[308]

6.3.2 Four Definitions of 'Culture'

- 'Culture is information that is acquired from other individuals via social transmission mechanisms such as imitation, teaching, or language… where information is intended as a broad term to refer to what social scientists and lay people might call knowledge, beliefs, attitudes, norms, preferences, and skills, all of which may be acquired from other individuals via social transmission and consequently shared across social groups'.[309]
- 'Culture is information that is acquired from other individuals via social transmission mechanisms such as imitation, teaching, or language… where information is intended as a broad term to refer to what social scientists and lay people might call knowledge, beliefs, attitudes, norms, preferences, and skills, all of which may be acquired from other individuals via social transmission and consequently shared across social groups'.[310]
- 'Culture is (mostly) information stored in human brains, and gets transmitted from brain to brain by way of a variety of social learning processes'.[311]
- 'In 1952 Alfred Krober and Clyde Kluck John researched 164 prior definitions pertaining to all cultures into one, as follows: Culture is a product; is historical; includes ideas, patterns and values; is selective; is learned; is based upon symbols; and is an abstraction from behaviour and the productions of behaviour'.[312]

'Culture' according to cultural evolutionists is *any information shared between organisms transmitted between and stored in brains, sometimes expressed through behaviour*.[313]

6.3.3 Cultural Transmission *v.* Genetic Transmission

[308] Churchland, *Braintrust*, 20.
[309] Mesoudi, *Cultural Evolution*, 3.
[310] Mesoudi, *Cultural Evolution*, 3.
[311] Ibid., 61.
[312] E. O. Wilson, *Consilience*, 142.
[313] Lewens, 'Cultural Evolution: Integration and Skepticism', 12.

> *Whereas genetic information is stored in sequences of DNA base pairs, culturally transmitted information is stored in the brain as patterns of neural connections and whereas genetic information is expressed as proteins and ultimately physical structures such as limbs and eyes, culturally acquired information is expressed in the form of behaviour, speech, artefacts and institutions.*[314]

If this is correct it follows we must be able to read off these neural patterns the content of the relevant cultural idea, though as yet ideas as to how we might do this are sorely lacking. Furthermore a point of interchange between these levels of selection is still required. This requirement is described below, followed by a clear example of such coevolution in relation to lactose intolerance:

> Genes prescribe epigenetic rules, which are the regularities of sensory perception and mental development that animate and channel the acquisition of culture. Culture helps to determine which of the prescribing genes survive and multiply from one generation to the next. Successful new genes alter the epigenetic rules of populations. The altered epigenetic rules change the direction and effectiveness of the channels of cultural acquisition.[315]

So we can see that genes via epigenetics have a more immediate capacity to reflect environmental changes than previously thought. Lactose intolerance is an apt example of this:

> Lactose intolerance due to environmental factors is one of the best documented examples of gene culture co evolution, in which cultures shape the selection of genes as much as the reverse.[316]

Boyd and Richerson, pioneers in this field, argue that the evolutionary process for culture has different properties to that of genes. 'Culture would never have evolved unless it could do things genes can't'.[317] If culture cannot be reduced the gene level, cultural evolutionists still have to show cultural evolution is genuinely Darwinian; thus Boyd and Richerson[318]

[314] Mesoudi, *Cultural Evolution*, 3.
[315] E. O. Wilson, *Consilience*, 171.
[316] David Sloan Wilson, *Does Altruism Exist?*, 64.
[317] Boyd and Richerson, *Not by Genes Alone*, 7.
[318] Ibid.

used population models (borrowed from evolutionary geneticists such as Fisher),[319] to determine how cultural transmission might operate, while more recently, (sliding from analogues to identities), Mesoudi has reduced the biological connection further, arguing that cultural change has all the essential, original Darwinian facets **of** variation,[320] competition[321] and inheritance,[322] and so the resources to account for convergence, adaptation and maladaptation.[323] Boyd and Richerson arrive at the same conclusion: 'the logic of natural selection applies to culturally transmitted variation every bit as much as it applies to genetic variation'.[324]

Critics point to the micro-level processes that give rise to Darwinian evolution. Micro-evolutionary discoveries include genetic inheritance being particulate, non-Lamarckian, and mutations being blind to selection. Thus Gould and John Maynard Smith, among others, reject cultural evolution because it does not seem to be Darwinian. For example, in cultural evolution, variation can occur in a 'directed' Lamarckian manner known as guided variation.[325] Another point of contention centres on the fact that cultural evolution takes a generous position on the issue of adaptation.[326] All in all, cultural evolutionists have serious biological critics, though this criticism has not stopped them from pursuing their research project.

> Stephen Jay Gould and evolutionary biologist Richard Lewontin have convinced many people, including many social scientists, that adaptive explanations are usually unjustified. Their position is that many features of organisms are historical accidents or side effects of adaptive changes in other characters, and that one must be extremely cautious in invoking adaptive explanations. *We couldn't agree less.*[327]

[319] R. A. Fisher, *The Genetical Theory of Natural Selection* (Oxford, England: Clarendon Press, 1930).
[320] Mesoudi, *Cultural Evolution*, 29.
[321] Ibid., 31.
[322] Ibid., 34.
[323] Ibid., 35.
[324] Boyd and Richerson, *Not by Genes Alone*, 76.
[325] Mesoudi, *Cultural Evolution*, 56. Mesoudi, of course, argues that the fact that cultural evolution is not Neo-Darwinian is irrelevant, because it is still Darwinian in the original formation
[326] Boyd and Richerson, *Not by Genes Alone*, 102.
[327] Ibid.

6.3.4 Applicability

Population models ('bookkeeping') work on the premise of a collection of individuals with certain traits (in this case cultural traits rather than biological ones) undergoing change, according to a specified set of processes or rules.[328] These acting processes are referred to as 'transmission':

> In cultural evolution inheritance takes the form of cultural transmission, the process by which information (e.g., knowledge, skills, or beliefs) is passed from individual to individual via social learning.[329]

Methodologies of transmission include cultural variation, cultural selection, cultural drift, natural selection and migration (diffusion).

According to Jeremy Kendal and Kevin Laland the techniques of cultural evolution can be used to predict how cultures change,[330] though thus far this predictive component has been a matter of retrodiction. However, there is no reason why, if we had enough information about an environment, we could not use cultural selection in a future orientated manner:

> Cavalli-Sforza and Feldman define cultural selection as a Darwinian process by which particular memes increase or decrease in frequency due to their differential probability of being adopted by other individuals. In contrast for these authors natural selection refers to the differential survival of individuals expressing different types of memes. For instance, the spread of contraceptive use through cultural selection processes could alter natural selection pressures induced by sexually transmitted diseases. Therefore the analysis is based on the explicit assumption that cultural traits evolve by Darwinian selection processes, whereby individuals can be selected purely on the basis of their memes.[331]

[328] Ibid., 57.
[329] Ibid., 3.
[330] J. R. Kendal and K. N. Laland, 'Mathematical Models for Memetics', *Journal of Memetics* 4 (2000), http://cfpm.org/jom-emit/2000/vol4/kendal_jr&laland_kn.html.
[331] Ibid.

In applying such evolutionary processes to the data set, we would expect a certain kind of 'evolutionary' result. The ultimate test then is whether the process modelling really matches the reality**.** It is this matching (or lack of it) that will determine how seriously we take the strictness-to-gene argument; a point not lost on cultural evolutionists who are flexible regarding the processes of cultural transmission as long as they accurately reflect cultural change in a population.[332] For Mesoudi the task is to generate quantitative predictions concerning the macro-evolutionary consequences of different micro-evolutionary processes, and to develop some way to identify such macro-evolutionary patterns; for this way we can test such quantitative predictions.[333]

The basic steps of Darwinian analysis used in cultural evolution are as follows:[334]

- Draw up a model of the life history of individuals;
- Fit an individual level model of the cultural (and genetic, if relevant) transmission process to the life history;
- Decide which cultural (and genetic) variants to consider;
- Fit an individual level model of the ecological effects to the life history and to the variants;
- Scale up by embedding the individual-level processes in a population and extend over time by iterating the one-generation model generation and generation.

Alongside content dependant psychological bias one would expect the more solid neural functions to be relevant to the transmission parameters of culture. Henrich, Boyd, and Richerson theorise memetic/cultural fitness (*v.* genetic fitness) by utilising a replicator dynamic equation for a specific rumour about a demon-fish; a rumour which if adopted would prevent a person from fishing at night when, in fact, that is the best time to fish.[335] This model offers a model for determining frequency of fitness.

$$\Delta p = p(1-p)[X-Z]$$

[332] Henrich, Boyd, and Richerson, 'Five Misunderstandings', 124.
[333] Ibid., 110.
[334] Boyd and Richerson, *Not by Genes Alone*, 97.
[335] Henrich, Boyd and Richerson, 'Five Misunderstandings', 126.

ΔP = frequency of individuals who adopt the culture, in that it modifies their behaviour.

X is strength of content bias (the psychological module processing or affected by the introduced culture—its genetic fitness)

Z is the cost to an individual who acts on the cultural information in terms of passing it on to another (its cultural fitness)

p is an individual.

If X is significant this raises the probability of cultural adoption, but this probability must be set against the influence of Z in terms of reducing the likelihood of survival and thus of the passing on of this culture in the long run.

Chapter 7: Cognitive and Behavioural Neuroscience—Part 1: Research Limitations and Ambiguities

7.1 Scanning for Moral Circuitry

To begin, it is important to note the distinction made in the title of this chapter. This enquiry concerns *behavioural and cognitive* neuroscience, where this identifies the variety of roles that are performed by multiple brain regions in producing an overall phenomenon[336] (e.g. a behaviour, a memory, a moral capacity). This is sufficient for the purposes of this study, which is concerned with cognitive decision making and the production of behaviour. The concern here is not with that *molecular* neuroscience which looks into the nature of these operations themselves.[337] The models discussed here are pitched at a *mechanistic level of explanation*. This area of neuroscience involves the study of the brain's regions by fMRI scan, and it has had a major impact on philosophy:

> No doubt when historians of science look back on the first decade of the twenty-first century, they will dub it 'The Age of the fMRI'… by one estimate, an average of eight peer-reviewed articles employing fMRI were published per day in 2007. So perhaps it was inevitable that empirically minded philosophers would take some of these fMRI studies to have profound implications for philosophy.[338]

At the very least, brain imaging has offered a great deal of scientific information about how the human brain functions, and is the best observational source currently available**.** What it has not shown – despite Michael S. Gazzaniga's claim – is that 'arguments that have raged for centuries about the nature of moral decisions and their sameness or difference are now quickly and distinctly resolved with modern brain imaging'.[339] Joshua Greene's formulation is more modest, better aligning with the present state of inquiry, and opens the door to a *relationship of manipulability*:

---

[336] Bechtel, 'Molecules, Systems, and Behaviour', 20.
[337] Ibid., 34.
[338] S. Berker, 'The Normative Insignificance of Neuroscience', *Philosophy and Public Affairs* 37 (2009): 293.
[339] Michael S. Gazzaniga, *The Ethical Brain: The Science of Our Moral Dilemmas* (New York: Harper Perennial, 2006), 167.

> Our current neuroscientific understanding of moral judgment is rather crude, conceptualized at the level of gross anatomical brain regions and psychological processes familiar from introspection. But, for all our ignorance, the physical basis of moral judgment is no longer a complete mystery**.** We've not only identified brain regions that are involved in moral judgment, *but have begun to carve the moral brain at its functional joints*.[340]

Normally what happens is that a 'moral sector' of the brain is identified, and named as such, through imaging. Psychologists and philosophers then use these images to support or deny certain interpretative claims about morality – claims, for instance, about morality's relationship to reason and emotion, and of the ordering of that relationship. At this level psychologists utilise cognitive theories to draw conclusions about how the brain arrives at X directives for behaviour, given Y stimuli, in a way not forthcoming at the level of molecular neuroscience.

So what does an fMRI scan actually do? It measures contrasts in levels of blood oxygenation in various areas of the brain (i.e. how much oxygen there is in the scanned area of the brain over one period of time compared to the next). This technique provides maps through the time of activated neural networks, though there are limitations. For instance, questions such 'can neurons still be activated without oxygen flow and does this impact consciousness?' remain unanswered, and will not be considered further in this study. Another more relevant issue with fMRI reports is that areas of the brain that are reflected as lighting up with moral problem solving often do so with non-moral problems as well.[341] This has led Churchland to suggest a 'single continuum' here:

> That social and moral behaviour are part of single continuum is modestly supported by neuroscientific data showing that whether a subject sees a merely social event or a conventionally 'moral' event, the same regions of the prefrontal cortex show increased activity.[342]

---

[340] Joshua Greene, 'The Cognitive Neuroscience of Moral Judgment', in *The Cognitive Neurosciences*, 5th ed., ed. M. S. Gazzaniga (Cambridge: MIT Press, 2009), 26.
[341] M. D. Hauser, *Moral Minds*, 241.
[342] Churchland, *Braintrust*, 60.

This suggests that best utilising fMRI findings requires extensive cross-field integration with psychology, evolutionary biology, anthropology and cultural evolution.[343]

While fMRI scanning fails itself to pierce through to the level of neural network connections, this does not mean the information it provides is not useful:

> fMRI does not reveal what is going on at the micro level of neurons and their networks. Without that, getting to the bottom of the operations and business of the PFC is problematic. Nevertheless to repeat fMRI is a wonderfully important technique for studying the brain organisation in humans, allowing us to see, for example, where there are differences between the brains of psychopaths and the brains of control subjects.[344]
>
> [e.g. Blair 2007 showed psychopaths have amygdala dysfunction (fMRI) reduced activity in moral dilemmas][345]

An external limitation of fMRI scanning lies in the experimental context itself. This context may bias the subject's brain operations when compared to those everyday contexts in which moral judgments are typically made. Indeed, all kinds of varying environmental influences may affect the moral decision under study:

> Ecological validity is especially relevant for moral cognition studies, because moral cognition depends strongly on situational and cultural context. The experimental constraints that are imposed by behavioural and functional imaging studies might have an important impact on performance of moral cognition tasks. Some people might feel uncomfortable disclosing their opinions about sensitive issues, providing socially desirable answers instead.[346]

Further, their cultural background may influence the type of response they give to certain moral dilemmas:

---

[343] J. Moll et al., 'The Neural Basis of Human Moral Cognition', *Nature Reviews Neuroscience* 6 (2005): 799.
[344] Churchland, *Braintrust*, 125.
[345] James R. Blair, 'The Amygdala and Ventromedial Prefrontal Cortex in Morality and Psychopathy', *Trends in Cognitive Science* 11 (2007): 387–392.
[346] Moll et al., 'Neural Basis', 803.

> … inferring cognitive and neural mechanisms from behaviours can be misleading, especially when cultural and situational factors are involved**.** For instance, Westerners and East Asians differ in categorization strategies when making causal attributions and predictions, and moral values and social preferences are shaped by cultural codification.[347]

These same objections apply to other forms of brain scanning, such as electroencephalography (EEG), magnetoencephalography (MEG), positron emission tomography (PET), and single photon emission computed tomography (SPECT).

At this point, the sceptical philosopher might ask why, given all these potential issues, we rely much at all on such inquiries? Greene replies with a smart and honest defence of the usefulness of fMRI scanning:

> First, studying the neuroscience of moral judgment adds value to neuroscience. Second, it adds value to both neuroscience and psychology by furthering their integration. Third, neuroscientific tests of psychological theories can be especially powerful because neuroscientific results, unlike behavioural results, cannot be predicted by common sense. (Neuroscientific results are rarely met with 'duh'.) Fourth, understanding the neuroscience of moral judgment may also play a valuable role in dispelling mind-body dualism (Greene, in press) a worthwhile philosophical endeavour, but not one that is likely to challenge the views of contemporary ethicists. In short, studying the neuroscience of moral judgment is valuable for several reasons that have little to do with normative ethics. Thus, if my fMRI studies of moral judgment, taken in isolation, fail to have definitive moral implications, this may be because having moral implications is not their most immediate purpose.[348]

7.2 Redefining the Subject of the Moral

[347] Ibid., 804.
[348] Joshua Greene, 'Notes on "The Normative Insignificance of Neuroscience" by Selim Berker', accessed 1 March 2014, http://www.wjh.harvard.edu/~jgreene/GreeneWJH/Greene-Notes-on-Berker-Nov10.pdf, 7.

Gabriel Abend, associate professor of sociology at New York university, is critical of the limited scope of neuroscientific and psychological investigations of morality. He challenges the 'moral judgement' focus which is at the 'methodological heart' of most experimental studies. Abend points out that what is missing in this focus, and so these analyses, is the role of *narratives* in determining moral judgement. Abend points out that researchers focus on moral judgements as a matter of *what to do (here, in this situation, now)*, whereas just as important, even more so, is the matter of persons deciding (or finding) *how to be*. This 'moral self-conception or identity, he says, is missed by talk about, right, wrong, permissible and forbidden'.[349]

A connected point of Abend's concerning the explanatory power of fMRI is that of the narrow causal sequence between stimuli and the 'moral' response. Abend points out that experimenters often simply assume that given a specific stimuli the causal response to it is the result of that immediate stimuli when in fact it may not be. A neuroscientist may present a trolley scenario as stimuli ('Would you push a man off a bridge to save three below?') and take it that if the recipient of the question says 'Yes', then they are making a utilitarian moral decision. However, a person's moral decision here may not depend on the immediate stimuli presented, but may instead be caused by a number of prior incidents and personal moral reflections specific to their life.

Abend's 'characteristics of morality' show us how much can be involved in the formulation of our morality, and why we need to be careful in making narrow causal association:

- A moral judgment is made by and is an attribute of one individual.
- It's made in response to a specific stimulus.
- The stimulus is an imaginary situation and a question about it.
- The judgment is about an action (rather than, say, a person or state of affairs).
- A moral judgment is a statement (indicative mood).
- It is in essence an utterance or speech act (even if not in fact uttered).
- It makes use of 'thin' ethical concepts only (okay, appropriate, permissible, acceptable, wrong, etc.).
- It's fixed, settled, verdict-like.

[349] Ibid., 172.

- It's clear (not conceptually or semantically muddled, incoherent, etc.).
- It's made at a specific, precise, discrete point in time.

Four additional features do not apply to the whole class of moral judgments, but only to a subclass—namely, judgments made in response to dilemmas (as in the Trolley case):

- The imaginary situation is unrealistic.
- It presents a choice between (usually two) courses of action or between action and inaction.
- It's about a future action.
- The subject of the choice is either an imaginary person or the experimental subject herself (in which case she is asked to imagine what she would do, or what would be appropriate for her to do, if she found herself in such a situation).

Many of these points relate to the issues discussed in scanning for moral circuitry; and certainly we should pay attention to Abend's points in two particular cases. First, we must be careful not to deploy a false reductionism through absolute language based on clearly limited experiments. (The methodological discussion on the MF are especially relevant here.) Second, we should appreciate that these experiments fall short of fully capturing the experience of moral decision, and that this experience may have causal ramifications.

This does not mean that we should not pay great attention to theories derived from experiments that might reveal useful causal information on aspects of our decision making. Even when there are gaps in our knowledge, we can still produce a mechanistic sketch and relationship of manipulability useful for policy construction and in assessing the effectiveness of any ethics caught up in or presupposed by that policy design**.** We simply need to be careful about what it is the science is actually doing and describing. Carefully then, let us move forward and consider what cognitive neuroscience has to offer here.

Chapter 8: Cognitive and Behavioural Neuroscience—Part 2: Decision Making

8.1 Introduction and Terminology

> *Your brain is organised to take pleasure in company and to find social exclusion painful. Your brain is organized to feel if your child is maimed or your mate is assaulted. If your brain were organized like that of a turtle, you would spend years alone, entirely contented. Nor would you care if a neighbouring turtle got made into soup or baby turtles got eaten by seagulls.*[350]

The interest in neuroscience from a decision-making perspective has dramatically increased over the last decade, based primarily on new technologies and a general belief in the reductive explanatory capacities of modern science. New researches using fMRI studies are starting to reveal the causal mechanisms behind what was formerly 'the impregnable organ'. While very few have looked at the neuroscience in light of concerns with environmental decision making, let alone at such decision-making in the Anthropocene, there has been significant research into various elements of economic and moral decision making that can be leveraged for such purposes. To do this we need to address the following:

- How does the brain value various ideas and objects in reality, and how does it make decisions with respect to them?
- Can a study of the brain's operations tell us anything about the potential success or failure of the adoption of specific (environmental) ethics in a policy framework, or provide any other practical information on decision making that may be useful in developing a response to the policy challenge?

Before we do this, some basic terminology must first be clarified. The term 'brain' is used to refer to the ontologically objective physical organ located within the skull, while the term 'mind' is used to refer to the 'brain' *and* conscious experience. As I have argued earlier, brains are ontologically objective and give rise to an ontologically subjective

[350] Churchland, *Touching a Nerve*, 104.

experience ('mind').[351] Now it is (at present) true that we do not know how to understand the 'emergence of consciousness from matter',[352] (the mind-body problem). This has led some to argue that while:

> Neurologists of the future may map brain activity so precisely that they will someday be able to read out what a person is consciously experiencing from that map—but the map will never be conscious experience. It can help explain a conscious experience, but only a person can understand it. Nature can only be explained; humanity can be understood, and understanding is a matter of meaning.[353]

This kind of argument has popped up in just about every critical philosophical account of neuroscience, but by itself it does not mean that there are not pragmatic benefits that may be obtained from such mapping – indeed benefits that presuppose and enhance our self-understanding. To treat this possibility fairly it would be wise, in the first instance, to adopt neuroscientists' definitions of morality:[354]

- 'Moral emotions' refer to emotions that are linked to the interest or welfare of other people, or society as a whole.[355] More technically, they are also defined as 'the co-activation of neural representations of social cues, event knowledge, and emotion … present regardless of whether the processing is conscious or unconscious'.[356] Jorge Moll, a leading neuroscientist in this area, has divided moral emotions into several categories: guilt, shame, embarrassment, pride, indignation/anger, contempt/disgust, pity/compassion, awe/elevation and gratitude. One key point about the moral emotions described here is their automaticity upon witnessed violation, in contrast to laborious, deductive reasoning.[357]

[351] Uttal, *New Phrenology*, 35.
[352] C. McGinn, 'Can We Solve the Mind–Body Problem?', in *The Mind–Body Problem: A Guide to the Current Debate*, ed. R. Warner and T. Szubka (Oxford, England: Blackwell, 1994), 114.
[353] T. de Zengotita, 'Ethics and the Limits of Evolutionary Psychology', *The Hedgehog Review* 15, no. 1 (2013): 34.
[354] Unless otherwise referenced, definitions are used here from Moll et al., 'Neural Basis'.
[355] Haidt 2003b.
[356] C. A. Hynes, 'Morality, Inhibition, and Propositional Content', in Sinnott-Armstrong, *Moral Psychology*, 3:25.
[357] Moll, 'Neural Correlates', 2730.

- ‘Moral judgement’ refers to a type of evaluative judgement that is based on assessments of the adequacy of one’s own and others’ behaviours according to socially shaped ideas of right and wrong, which we may refer to as prohibitions (innate or not).
- ‘Moral reasoning’ refers to the thought mechanism through which moral judgements are attained. Such reasoning may be only partly conscious.
- ‘Moral values’ refers to culturally shaped concepts and attitudes that code for personal and societal preferences and standards which take the form of prohibitions.[358]
- ‘Moral dilemmas’ are situations in which every possible course of action breaches some otherwise binding moral principle.[359] Tasks that bring morality under experimental scrutiny include presenting visual sentences or pictures or using scales and questionnaires to assess moral behaviour from a clinical point of view. Such dilemmas are utilised in attempt to expose the underlying processes and causality of the human brain *in situ*.

8.2 Neural Connectivity

As the *Oxford Dictionary* succinctly says; the brain is an ‘organ of soft nervous tissue contained in the skull of vertebrates, functioning as the coordinating centre of sensation and intellectual and nervous activity’. This may be obvious, but as soon as we probe further simplicity dissipates into complexity and scientific uncertainty, even on the level of such basic statements as ‘the brain is functional’.[360] In this section, I outline what we think we know about the brain, in terms of its structure and functionality as it relates to that understanding of morality set out above.

[358] L. Young and J. Dungan, ‘Where in the Brain is Morality? Everywhere and Maybe Nowhere’, *Society for Neuroscience* 7 (2012), 1–10.
[359] J. J. Thomson, ‘The Trolley Problem’, *Yale Law Journal* 94 (1985): 1395–1415.
[360] A. Sandberg and N. Bostrom, ‘Whole Brain Emulation: A Roadmap’, Technical Report #2008‐3 (Oxford, England: Future of Humanity Institute, Oxford University, 2008), www.fhi.ox.ac.uk/reports/2008‐3.pdf.

The simple reductionist lens tells us that brain is the most fundamental ground and source of our moral and ethical cognition. It is the brain that facilitates all incoming stimuli from the body and environment and produces our responses.

According to medical literature, the primary functional unit of the brain is the *neuron*.[361] A neuron is simply a nerve cell, a transmitter of electrochemical messages. There are approximately 100 billion neurons in a human brain, and each one is connected to between 1,000 and 10,000 other neurons. The point of intersection between two neurons is referred to as a *synapse*, and is where *communication* between the neurons occurs, for neurons can either be 'on' or 'off'.[362] In some cases, this switch is controlled by genes. It was Richard Canton in the nineteenth century who discovered that a brain responds to and operates via electrical current, but one should also recognise the brain is dually a biological process:[363] Neural synapses 'are energised by the oxidation of glucose—a bio chemical process'.[364]

The central question in the study of cognitive neuroscience is how coordinated, purposeful behaviour arises from the activity of neurons.[365] Medical research has shown how neurons connected to form certain structures; some with discrete cognitive functions. When two or more of these structures interact they are referred to as *circuits*, which pass information between each other in a back and forth manner.[366] The brain is localised/compartmentalised, and neuroscience has identified specific functional areas such as sensory and motor regions: 'With regard to sensory and motor functions, it has become indisputable that the brain is not equipotential or homogeneous, but is made up of a cluster of relative specialled regions'.[367] Some of these areas are located in physically unique and identifiable sections of the brain, others have been identified through patterns of neural activity spanning multiple regions. Although localised regions exist in the brain for such functions as motor control, brain functions also occur as a 'mass action' and unfortunately

[361] Nobel Prize winner Santiago Ramon y Cajal (1880) discovered that the brain is not mass tissue but rather discrete, connected neurons.
[362] V. S. Ramachandran, *The Tell-Tale Brain: Unlocking the Mystery of Human Nature* (London: Windmill Books, 2012), 13.
[363] 'The place where a signal passes from the neuron to another cell'. http://www.ninds.nih.gov/disorders/brain_basics/know_your_brain.htm
[364] Uttal, *New Phrenology*, 70.
[365] E. K. Miller and J. D. Cohen, 'An Integrative Theory of Prefrontal Cortex Function', *Annual Review of Neuroscience* 24 (2001): 167, http://search.proquest.com.ezproxy.une.edu.au/docview/198920013?accountid=17227.
[366] Ramachandran, *Tell-Tale Brain*, 15.
[367] Ibid., 11.

for explanatory purposes, 'the more complex the psychological process, the less likely it is that narrowly circumscribed region uniquely associated with that process will be found'.[368] The neuroscientific research into decision making explored in the next section stands as testament to this.

Even though we know a certain area is responsible for motor control, the process (for example) of moving an arm often involves stimulation of multiple areas of the brain. Not surprisingly then, neuroimaging shows that *moral judgement* involves a wide variety of brain areas, some of which are argued to be extremely ancient.[369] Today's leading framework involves a *network*-based concept of the brain, where individual regions perform various functions.[370] Such neural networks are sought after to explain the origin of thought and behaviour. When it comes to morality there is clear localisation, with moral emotions located in specific areas and moral cognition in other areas.[371]

When we talk about the brain making a decision utilising a specific set of areas or 'neural pathways' we are discussing a *brain function.*[372] Haidt gives a succinct analysis of such functionality:

- It is computational as either 'neuron'[373] or 'neuron assemblies'[374] (either way brain function is a microscopic neural phenomena).
- Due to computational tractability, reduction to the mapping of neural nets [brain functions] is likely impossible.[375] (However areas such as motor control have been

[368] Ibid., 13.
[369] de Waal, *Primates & Philosophers*, 56.
[370] F. Schirmann, 'Invoking the Brain in Studying Morality: A Theoretical and Historical Perspective on the Neuroscience of Morality', *Theory & Psychology* 23, no. 4 (2013): 297.
[371] Aside from theorising about morality by way of brain imaging and mapping, there are those (e.g. Churchland) who specify some kind of biological substance that influences the brain and morality, such as serotonin or oxytocin.
[372] I have chosen to use the word brain 'function'. However, in philosophy, brain 'process' was used to identity theory (see Feigl 1967; Smart 1991; Place 2004). Others have used 'brain state' (e.g. Lewis 1966; Armstrong 1968). I do not like the static nature of this term, and prefer 'function', since it is the most appropriate for a neuroscience analysis of the brain. For an excellent account of terminology in this area, see Brown 2006.
[373] The 'Neuron Doctrine' is outlined in H. B. Barlow, 'The Neuron Doctrine in Perception', in *The Cognitive Neurosciences*, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press, 1995).
[374] Singer argues that the functional unit of the brain is neuronal assemblies, as opposed to the neuron. Singer 2000.
[375] Uttal, *New Phrenology,*13.

identified and can be stimulated consistently with electricity to, for example, cure Parkinson's disease.)[376] (Parkinson's hasn't itself been cured though).

- Consciously experienced, or not, 'most of cognition occurs automatically, and outside of consciousness'.[377]

Popular but controversial neuroscientist Chris Firth has suggested that our brains operate and learn without conscious awareness. He refers to this as 'learning by association'. The general idea is that synapses communicate with each other via neurotransmitters. Dopamine is one type of neurotransmitter. Wolfram Schultz's tests on monkeys that had to press buttons to receive a reward showed that the dopamine neurotransmitter activated when the monkey pressed the button and received the reward.[378] The repetition of stimuli leading up to the button press (a light activated first by the experimenter) showed that the neurotransmitter was activated *prior* to the actual reward being received. When the reward did not eventuate, the neurotransmitter weakened. The idea is that with us too our brains, often independently of consciousness, learn in this way about the outside world. In this account, it is just this process that creates what we refer to as *value,* for what is valued is anything that occurs prior to the reward being received.[379] Dopamine as the mediating reward has largely been abandoned in favour of reward prediction error theories.

We might only consciously think of the reward itself, but our brain's determination of value revolves around the success and failure implicit in the synapse dopamine process. This unconscious representational process is facilitated through a biological process and lets us see how the brain maps and learns about the outside world. This can be represented as a process on a computer via the temporal difference algorithm.[380] It is also important to note that this habitual, trial-and-error-based decision making, can be shaped in content by culture (i.e. 'fearing spiders').

[376] Andres Lozano, 'Parkinson's, Depression and the Switch that Might Turn Them Off', January 2013 at TEDxCaltech
[377] Haidt, 'Emotional Dog', 830.
[378] W. Schultz, 'Reward Signalling by Dopamine Neurons', *Neuroscientist* 7, no. 4 (2001): 293–302.
[379] C. D. Firth, *Making up the Mind: How the Brain Creates Our Mental World* (Oxford: Blackwell, 2007), 95.
[380] W. Schultz, P. Dayan, and P. R. Montague, 'A Neural Substrate of Prediction and Reward', *Science* 275, no. 5306 (1997): 1593–1599.

There is also a 'special class' of neurons referred to as *mirror neurons*, discovered over 25 years ago in the ventral premotor region F5 of the macaque monkey.[381] These neurons are special because they fire not only when you perform an action, but also when you watch *someone else* perform the same action.[382] This is what turns culture into a new genome:

> By hyper developing the mirror neuron system, evolution in effect turned culture into the new genome. Armed with culture, humans could adapt to hostile new environments and figure out how to exploit formerly inaccessible or poisonous food sources in just one or two generations, instead of hundreds or thousands of generations such adaptations would have taken to accomplish through genetic evolution.[383]

This, then, is effectively a value system that works in interaction with others, and it seems plausible that it is one that has played a significant role in the development of our morality:

> The evolution of intersubjectivity and empathy are due as much to the development and complexification of mirror-neuron systems as to the emergence of natural language.[384]

And:

> Mirror neurons appear to mediate our understanding of emotional states via imitation, allowing the translation of an observed action (such as a facial expression) into its internally felt emotional significance. This translation appeared to be absent in autism.[385] [386]

---

[381] di Pellegrino et al., 'Understanding Motor Events: A Neurophysiological Study', *Experimental Brain Research* 91 (1992): 176.
[382]Ramachandran, *Tell-Tale Brain,* 23.
[383] Ibid.
[384] C. Laughlin and J. Loubser, 'Neurognosis, the Development of Neural Models, and the Study of the Ancient Mind', *Time and Mind: The Journal of Archaeology, Consciousness and Culture* 3, no. 2 (2010): 146.
[385] U. Goswami, 'Neuroscience and Education: From Research to Practice?', *Nature Reviews Neuroscience* 7, no. 5 (2006): 412.
[386] Of course, there is still uncertainty surrounding the role of mirror neurons. Further studies are required to prove exactly what they are capable of, and if, for example, they allow us such capacities as visualisation, dreaming and imagination.

However, it should also be considered that there is still doubt as to what 'mirror neurons' really are, and if they exist in humans.

Now each person's brain is ontologically unique, in terms of genealogical composition,[387] space-time location, and the external stimuli it has encountered. Every neuron in a brain is also unique.[388] Thus each brain function is objectively unique. However, that all of our brains are unique, does not mean they are not functionally identical. Two cogs may be unique in size, shape and specific matter, but still perform the same function.

8.3 Neural Decision Making

8.3.1 Motivation

Anthony Landreth, in the context of policy concerns centred on drug addiction, proposes a mechanistic account of motivation. The idea is that the control systems in the brain supply the brain with goals, and set about achieving these through a set of adaptive sub goals.[389]

> Considering the enormous budget devoted to the war on drugs, the success rate of current treatments, and the toll addiction takes on people's private lives, any advance on understanding motivation mechanisms is attractive. Perhaps one a day a precise delivery system could be developed to work on the specific desire mechanisms underlying an addictive condition. Taking the notion a step further, one could imagine a future where people are given direct control over any number of their desires, enabled to turn down the volume on self-defeating forms of motivation, and turn up the volume on nobler motives.[390]

Landreth proposes various connective models and their component parts required for such a system to operate, and explores the known neuroscience of the brain to find a possible fit within the relationships of its regions. Evidence for a potential fit comes from both fMRI and behavioural studies. The most significant idea to emerge from such testing concerns

[387] Barlow, 'Neuron Doctrine', 4.
[388] Perry, 'Neurons: A Curious Collection of Shapes and Sizes', *BrainFacts.org*, last modified 22 August 2012, http://www.brainfacts.org/brain-basics/neuroanatomy/articles/2012/neurons-a-curious-collection-of-shapes-and-sizes/. Here, Perry references neuro-scientist Kristen Harris.
[389] Anthony Landreth, 'The Emerging Theory of Motivation', in Bickle, *Oxford Handbook*, 383.
[390] Ibid., 415.

value proxies, whereby the predictors of rewards in a dopamine system substitute for the reward itself.[391] This is thought to occur in the orbitofrontal cortex (OFC), and it may override habitual responses governed by the ventral striatum.[392]

Assuming eventual alignment between motivational theory and the actual workings of the brain, Landreth's hopes for motivational control do not seem unrealistic. Even the mere awareness of these systems may allow one to consciously manage one's desires and motivations by searching out stimuli that may act as a value proxy. Indeed the practical possibilities seem endless. But is it really this simple? It is worth taking a closer look at valuation and choice modelling.

8.3.2 Valuation and Choice Models

Table 8.1: Areas of the Brain

Frontal Lobe	Parietal Lobe	Temporal Lobe	Limbic Lobe	Sub-Cortical Structures
Ventromedial Prefrontal Cortex (VMPFC)	The Inferior Parietal Region	Superior Temporal Sulcus (STS)	Posterior Cingulate Cortex (PCC)	The Hippocampus
Dorsolateral Prefrontal Cortex (DLPFC)	The Temporoparietal Junction (TPJ)	The Anterior/Middle Temporalgyrus	The Insular Cortex	The Amygdala
Orbitofrontal Cortex (OFC)	Lateral Intraparietal Cortex		The Anterior Insular Cortex	The Thalamus
Anterior Cingulate Cortex (ACC)				The Striatum

Neuroscientists have tended to assess the way the evolved brain produces moral decisions rather than at (the possibility of) causation the other way. Far from assuming that moral decision x = y kind of brain state, they assume (often on the basis of evidence from pre-frontally damaged brains) that causation runs the other way.[393] The way the brain determines moral decisions presumably involves a form of valuation among alternatives, whether inclusive of consciousness or not. I now look at research on valuation in more detail.

[391] Ibid., 391.
[392] Ibid., 392.
[393] Schirmann, 'Invoking the Brain', 293.

Perhaps due to research funding opportunities, the neural modelling of decision making actually began in the field of economics. Today there is a specialised field for this endeavour called *neuroeconomics* that seeks to explain behavioural causes by understanding how the brain makes value-based decisions.[394] It is surmised that decision making is a matter of evaluation and choice (deliberation) whereby multiple inputs are considered, one is prioritised, then chosen, and then (circumstances permitting) executed. A recent study provides two algorithms representing this brain process.[395]

The first facilitates valuation (learning, storing, representing), while the second facilitates choice. Brought together they form a two-step process, each stage of which utilises particular areas of the brain. Crucially for this thesis, with its ultimate policy orientation, such knowledge has already proven useful in understanding aberrant decision making in addiction, psychiatric disorders, autism and Parkinson's disease.[396]

Valuation, the first component, gained notoriety with a study of capuchin monkey decision making, which demonstrated cognitive processes beyond learning by trial and error. Padoa-Schioppa and Assad showed that these monkeys evaluated the food pairings placed in front of them based on their quantity and quality. [397] Traditionally, this would be accounted for by a stimulus model in which evaluation takes place only in terms of prior association. However, a second experiment with novel food pairings, showed a consistent form of evaluation, leading the researchers to consider representation in value cognition. They then conducted a third experiment, which involved giving the monkeys different types of juice pairings and mapping their neurons in the OFC. The results revealed three neuron paths that fired in a specific sequence, representing three forms of value criteria: *offer value*, *chosen value*, and *taste neurons*. The key finding was that the neural pathways did not alter when the juice menu changed (an invariant result). This is only possible if each of these neuronal pathways deploy a common evaluative scale. This turned out to be the case.

[394] Milton Friedman, *Essays in Positive Economics* (Chicago: University of Chicago Press, 1953).
[395] J. Kable and P. Glimcher, 'The Neurobiology of Decision: Consensus and Controversy', *Neuron* 63 (2009): 733.
[396] Kable and Glimcher, 'Neurobiology', 743.
[397] C. Padoa-Schioppa and J. A. Assad, 'Neurons in the Orbitofrontal Cortex Encode Economic Value', *Nature* 441 (2006): 223–226.

Still, however interesting the finding of value transitivity, we might ask what are these subjective values? Are they simply subjective preferences, or are they related to the features of external objects? It has been argued that such values are, in fact, a measure of the difference between the subjective and the external.[398] What we have is a 3 part valuation process that codes information inputs (both subjective reflections and information provided by/from external objects), then the choice model (the next part) takes that information and makes a choice determined by the difference between the two (you get more of x than y).

The learning of subjective preferences that are weighed up in this process is well understood. In a separate study using fMRI studies of neurons in an area of the capuchin brain called the Striatum (which humans also have) researchers were able to develop a 'reinforcement learning algorithm' that could give the probability of the monkeys' future actions.[399] The basic idea is that these values (subjective preferences and object identities in the external world) are encoded in the striatal neurons, where (presumably) they are engaged by the choice module (Step 2).

Choice clearly concerns the selection and execution phase of decision making. Central to this process is retaining that which was valued prior to the eventual choice, and the anticipation of the consequences of such choices. It is no surprise then that areas of the brain that involve working memory are utilised, specifically those found in the frontal lobe, though areas in the parietal lobe are also involved. Several studies (most famously those concerning Phineas Gage) have shown human patients with PFC damage perform particularly poorly on working memory tasks. Although the encoding of subjective values is invariant, when accessed in terms of the choice step they become relative. Utilisation is determined by weighting them on a single scale, which occurs through a differential analysis of neuron firing rates specific to the subject values in a specific set of areas in the parietal lobe. The subject value that is chosen is the one that has a firing rate that exceeds a local threshold, triggering selection through activation of other brain areas.[400]

[398] Paul Glimcher, 'Choice: Towards a Standard Back-Pocket Model', in *Neuroeconomics: Decision Making and the Brain*, ed. P. Glimcher, C. Camerer, E. Fehr, and R. Poldrack, 503–521. New York: Academic Press, 2000. http://www.decisionsrus.com/documents/choice_towards-a-standard-back-pocket-model.pdf.
[399] K. Samejima et al., 'Representation of Action-Specific Reward Values in the Striatum', *Science* 310 (2005): 1337–1340.
[400] Paul Glimcher (2009: 511–12)

The discovered invariance suggests that the brain automatically creates a kind of common currency, in much the same way as CBA does.[401] In this sense, what has value, whether economic or moral, is determined by representations in the PFC under a common scale. This suggests that in many circumstances our moral decision making may turn out to be largely predictable, as the coding structure is genetic.

Let us consider how these firing neurons representing options are coded, and how one is determined relative to another so that a choice results. We know that neural activity causally determines which choice is made (through stochastic neural computations[402]). A recent finding by Louie, Khaw, and Glimcher[403] involving tests on human subjects, suggests that the neural mechanism of value coding critically influences stochastic choice behaviour, opening the way to a generalizable quantitative framework for examining context effects in decision making. This is because the 'neural representation of the value of an option is explicitly dependent on the value of other available alternatives'.[404]

Louie, Khaw, and Glimcher argue for the occurrence of *decisive normalisation* to explain how certain sensory systems make decisions between competing data. The idea is that 'the initial input driven activity of a neuron is divided by the summed activity of a large pool of neighbouring neurons'.[405] The authors adopted this approach and discovered it produced the same results as value-guided choice experimentation performed on monkeys and human beings. Under this model, the value of an option represented is calculated by the mean firing rate of neurons; the firing rate incorporates variability in terms of noise, which is added in the equation to the mean firing rate; and the choice made is the option with the maximum firing rate (Figure 8.1).[406]

[401] In order to determine whether benefits outweigh costs, it is desirable to attempt to express all benefits and costs in a common scale or denominator, so that they can be compared with each other, even when some benefits and costs are not traded on markets and hence have no established dollar values. (S. Kelman, 'Cost Benefit Analysis: An Ethical Critique', in Van de Veer and Pierce, *Environmental Ethics and Policy Book*, 329.)
[402] E. Nica and G. H. Popescu, 'Neuroeconomic Models of Decision Making', *Knowledge Horizons – Economics* 6, no. 1 (2014): 63–66.
[403] Louie, Khaw, and Glimcher, 'Normalization is a General Neural Mechanism for Context-Dependent Decision Making', *Psychological and Cognitive Sciences* 110, no.15 (2013): 6139–6144.
[404] Ibid., 6139.
[405] Ibid., 6139.
[406] Ibid., 6140.

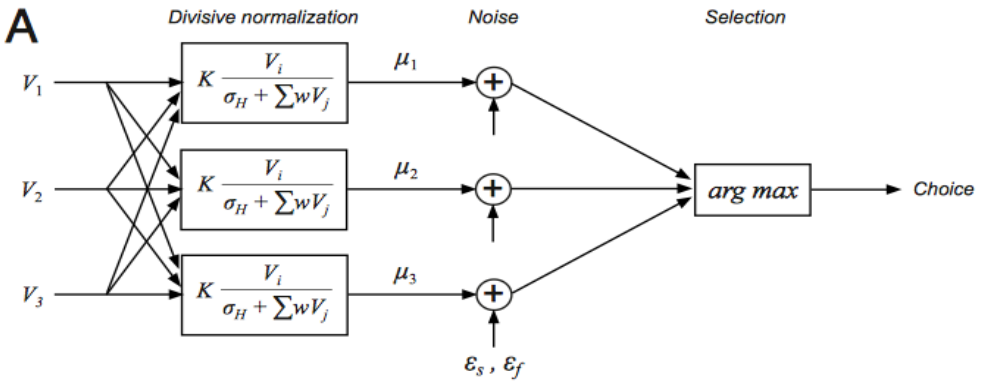

Fig. 1. Model of normalized value coding in stochastic choice. (*A*) Structure of the trinary-choice model. Value inputs (V_1,V_2, and V_3) were converted into mean firing rates (μ) representing each choice option and variability was introduced as noise terms added to the mean firing rates. In a given trial, the option with the maximum firing rate was designated as the chosen option.

Figure 8.1: Model of Normalised Value Coding in Stochastic Choice

Although there have been many behavioural studies on context-dependence, prior to this study it was generally held that choice was context-independent, so 'how a chooser decides between any two options should not depend on the number or quality of other options, a property known as the independence of irrelevant alternatives'.[407] The new evidence suggests that this conclusion may be false. (We should note that this study did not examine multi-attribute decision making involving the PFC. And presumably much of our moral and ethical decision making involves this part of the brain, more so than in the case of economic/reward-based decision making.)

At this juncture, let us look more deeply into emotion and risk taking as it influences decision making in the economic and risk-taking scenarios.[408]

8.3.3 Emotional Decision Making

> *Social problem solving is a messy practical business within an individual brain, where many interacting facts push, pull complete, and constrain the decision it settles on. Some constraints*

[407] Ibid., 6142.
[408] Ethical decision making in relation to the environment is just as likely, if not more so, to be made by these types of decisions rather than those with conscious consideration of refined ethical values.

take priority over others; some factors will be conscious, others not; some can be articulated, some not.[409]

A number of recent neuroscientific studies have measured the effect of 'incidental affective states',[410] such as mood or stress, on decision-making.[411]

Such studies have measured the effects of stress and mood across a broad range of scenarios. They may concern themselves with *incidental stress* (being stressed then making a choice), or *stressful choices* themselves.[412] Neuroeconomics is concerned primarily with habitual decision making in the Striatum, a sub-cortical structure, and there are many psychological studies here that are targeted at decision making under stress, though not many involve brain-imaging studies. For example, Porcelli and Delgado showed that stress intensified risk seeking when choosing between possible losses, but resulted in a risk-adverse approach when choosing between potential gains.[413] Youssef et al. showed that higher stress levels alters personal moral decision making, with a decreasing likelihood of people making utilitarian judgements.[414] Nguyen and Noussair found via face scanning software under a number of test conditions that positive emotions (so, not being stressed) correlated with greater risk taking.[415] Kimura et al. showed that stress exaggerates the tendency to discount future rewards in favour of smaller immediate rewards.[416] Harle et al. showed people are more likely to reject unfair offers when they are in a sad mood.[417]

[409] Churchland, *Braintrust*, 23.
[410] E. A. Phelps, K.M. Lempert, and P. Sokol-Hessner, 'Emotion and Decision Making: Multiple Modulatory Neural Circuits', *Annual Review Neuroscience* 37 (2014): 263–87.
[411] Such neuroscientific studies are already undertaken on safety in aviation decision making, and one can see how they might bear equally on environmental policy decision making. (For the former, see Mickaël Causse et al., 'The Effects of Emotion on Pilot Decision-Making: A Neuroergonomic Approach to Aviation Safety', *Transportation Research Part C: Emerging Technologies* 33 (2013): 272–281.
[412] Phelps, Lempert, and Sokol-Hessner, 'Emotion and Decision Making', 273.
[413] A. J. Porcelli and M. R. Delgado, Acute stress modulates risk taking in financial decision making. Psychol. Sci. 20 A. J. Porcelli and M. R. Delgado, 'Acute Stress Modulates Risk Taking in Financial Decision Making', *Psychological Science* 20 (2009): 278–83.
[414] F. F. Youssef et al., 'Stress Alters Personal Moral Decision Making', *Psychoneuroendocrinology* 37 (2012): 491–98.
[415] Y. Nguyen and C. N. Noussair, 'Risk Aversion and Emotions', *Pacific Economic Review* 1, no. 3 (2014): 296–312.
[416] K. Kimura et al., 'The Biological Effects of Acute Psychosocial Stress on Delay Discounting', *Psychoneuroendocrinology* 38, no. 10 (2013): 230–8.
[417] K. M. Harle et al., 'The Neural Mechanisms of Affect Infusion in Social ´Economic Decision-Making: A Mediating Role of the Anterior Insula', *NeuroImage* 61 (2012): 32–4.

Lighthall et al. (here with corresponding fMRI evidence) showed that under stress males are more risk adverse than females.[418]

But what, taken together, do such studies really tell us? That ideally people should be stress free when decision making? that, in dangerous circumstances, they should be sad males, and so forth? While such questions might ridiculous (*"Corporations and councils are full of sad risk-averse males! And look what happens anyway!"*) experimental evidence shows otherwise. It is therefore wrong to be too cynical here, as with our parenthetical commenter. For instance, resource companies deploy human resources (HR) testing to determine the 'right kind of people' for their work cultures, and there seems no good reason why (for example) the department of environment, or an environmental organisation of some other kind, should not do so as well.

An education in our natural tendencies in these areas would allow greater awareness of those biases that impact on our ability to reason effectively, and so let us reduce or counteract such biases. After all, we have posters at work that tell us how to lift heavy objects, because we are naturally inclined to do it the wrong way, and it helps when others observe our faults; so why not, when the information is precise enough, something similar when it comes to our decision making generally? In fact, we already deploy this capacity when we notice somebody is showing behavioural signs of stress or anxiety etc., so why not further this capacity as it aligns with specific organisational aims?[419]

There are studies into stress and anxiety effects on the brain that are concerned with goal-directed decision making involving the PFC. The PFC is particularly affected by stress, as stress impairs goal-directed performance, with the brain sometimes reverting to habitual automatic decision making under such conditions.[420] Having explored the automatic, and now emotional, coding of subjective value in terms of basic decision making, we turn now to the PFC and higher reasoning, exploring models that seek to integrate those 'head and heart' mechanisms behind our decisions.

[418] N. R. Lighthall et al., 'Gender Differences in Reward Replated Decision Processing under Stress', Social *Cognition Affect Neuroscience* 7 (2012): 476–84.
[419] Phelps, Lempert, and Sokol-Hessner, 'Emotion and Decision Making', 276.
[420] L. Schwabe and O. T. Wolf, 'Stress Prompts Habit Behaviour in Humans', *Journal of Neuroscience* 39, no. 22 (2009): 7191–98.

Chapter 9: Cognitive and Behavioural Neuroscience—Part 3: The Moral PFC?

9.1 The Frontal Lobe

Wood and Graftman offer a succinct breakdown of the PFC:

> The PFC can be divided into ventromedial and dorsolateral regions, each of which is associated with posterior and subcortical brain regions, the ventromedial PFC has reciprocal connections with brain regions that are associated with emotional processing (amygdala), memory (hippocampus) and higher-order sensory processing (temporal visual association areas), as well as with dorsolateral PFC. The dorsolateral PFC has reciprocal connections with brain regions that are associated with motor control (basal ganglia, premotor cortex, supplementary motor area), performance monitoring (cingulate cortex) and higher-order sensory processing (association areas, parietal cortex). The ventromedial PFC is well suited to support functions involving the integration of information about emotion, memory and environmental stimuli, and the dorsolateral PFC to support the regulation of behaviour and control of responses to environmental stimuli.[421]

Brain imaging experiments have shown that many areas of the brain are, or may be, involved in our making moral judgements,[422] but it is clear that the dorsolateral prefrontal cortex (DLPFC), the Ventromedial prefrontal cortex (VMPFC), and the Amygdala, are the most important.[423] The evolution of the PFC is intimately related to the emergence of human morality.[424] PFC neurons are able to fire over extended periods of time, and so across events over time, allowing for long-term goals and their satisfaction. Thus the PFC is important when it comes to making decisions, controlling our impulses, and attributing

[421] J. N. Wood and J. Grafman, 'Human Prefrontal Cortex: Processing and Representational Perspectives', *Nature Reviews Neuroscience* 4, no. 2 (2003): 140.
[422] Moll et al., 'Neural Basis', 804.
[423] Greene, *Moral Tribes*, 123.
[424] Ibid., 799.

goals and perceptions to others.[425] We also know that the PFC acts like a 'conductor of other regions of the brain, with pathways to emotional brain structures'[426] that presumably pass information to it.[427] To be more specific, the PFC is 'a collection of interconnected neocortical areas that sends and receives projections from virtually all cortical sensory systems, motor systems, and many subcortical structures'.[428] Miller and Cohen elucidate this with a famous metaphor:

> The function of the PFC can be likened to that of a switch operator in a system of railroad tracks. We can think of the brain as a set of tracks (pathways) connecting various origins (e.g. stimuli) to destinations (responses). The goal is to get the trains (activity carrying information) at each origin to their proper destination as efficiently as possible, avoiding any collisions. When the track is clear (i.e. a train can get from its origin to destination without risk of running into any others), then no intervention is needed (i.e. the behavior can be carried out automatically and will not rely on the PFC). However, if two trains must cross the same bit of track, then some coordination is needed to guide them safely to their destinations. Patterns of PFC activity can be thought of as a map that specifies which pattern of 'tracks' is needed to solve the task. In the brain, this is achieved by the biasing influence that patterns of PFC activity have on the flow of activity in other parts of the brain, guiding it along pathways responsible for task performance, just as activation of the color-control unit in the Stroop model biased processing in favor of the color-naming pathway. Note that this function need not be restricted to mappings from stimuli to responses but applies equally well to mappings involving internal states (e.g. thoughts, memories, emotions, etc.), either as 'origins' or 'destinations', or both. Thus, depending on their target of influence, we can think of representations in the PFC as attentional templates, retrieval cues, rules, or goals, depending on whether the biasing influences target sensory processes, internal processes, particular courses of action, or their intended outcomes.[429]

[425] Churchland, *Braintrust*, 208.
[426] Ibid., 119.
[427] Greene, *Moral Tribes*, 137.
[428] Miller and Cohen, 'Integrative Theory', 168.
[429] Ibid., 184.

Miller and Cohen describe the types of brain functions that result in certain behaviours by way of a division in the type of processing the brain performs.[430] *Bottom-up processing* is responsible for behaviours that are simple and automatic. They occur in response to specific stimuli, and take time and experience to build up. We do not have to 'think' in these cases. By contrast, *top-down processing*, performed by the PFC, takes into consideration a range of information (internal and external) to arrive at the appropriate behavioural response. 'The PFC is critical in situations when the mappings between sensory inputs, thoughts, and actions either are weakly established relative to other existing ones or are rapidly changing'.[431] The PFC facilitates complex behaviours, rather than immediate stimulus responses. While bottom-up processes reflect the most recent stimuli, the PFC is able to 'selectively maintain task-relevant information':[432]

> Neurons in the PFC are particularly able to fire over extended periods of time, and across events. This indicates that the PFC can maintain stimulus representations across time … enabling a subject to engage in behaviour to achieve long-term goals.[433]

Strong evidence for this comes from the Stroop task:

> In the Stroop task (Stroop 1935, MacLeod 1991), subjects either read words or name the color in which they are written. To perform this task, subjects must selectively attend to one attribute. This is especially so when naming the color of a conflict stimulus (e.g. the word GREEN displayed in red), because there is a strong prepotent tendency to read the word ('green'), which competes with the response to the color ('red'). This illustrates one of the most fundamental aspects of cognitive control and goal-directed behavior: the ability to select a weaker, task-relevant response (or source of information) in the face of competition from an otherwise stronger, but task-irrelevant one.[434]

[430] Ibid.
[431] Ibid.
[432] Ibid., 180.
[433] Wood and Grafman, 'Human Prefrontal Cortex', 140.
[434] Miller and Cohen, 'Integrative Theory', 169.

We also know, thanks to studies of persons with brain damage, what it is the PFC manages:

> A patient seems to lose all interest in his own future and he shows no moral compunctions of any kind. He may laugh at a funeral or urinate in public. The great paradox is that he seems normal in most respects: his language, his memory, and even his IQ are unaffected. Yet he has lost many of the most quintessential attributes that define human nature: ambition, empathy, foresight, a complex personality, a sense of morality.[435]

Testing by Mario Mendez has shown that patients suffering damage to the VMPFC (in the form of Frontotemporal Dementia) are more likely to make utilitarian judgements. A further study supported by Koenigs et al. found that patients with VMPFC lesions are significantly more likely to endorse utilitarian responses to moral dilemmas.[436] Eslinger and Damasio observed an absence of moral behaviour from patients with lesions, but (surprisingly) a normal moral response when presented with moral reasoning tasks.[437] Subsequent studies have further shown that damage to the frontal lobes results in a failure to take into account future consequences when coming to a decision.[438]

A further part of the frontal lobe is the DLPFC. fMRI studies by Greene (2004) suggest this area is involved in cognitive control and problem-solving. It has also been argued that the DLPFC has a significant role in the judgment of responsibility for crimes, and its punishment, from a third-party perspective.[439]

Another key area in the frontal lobe is the Anterior Cingulate Cortex (ACC), which is involved in error detection and is activated when subjects produce a utilitarian response.[440]

9.2 Representation *v.* Processing

[435] Ramachandran, *Tell-Tale Brain*, 21.
[436] M. Koenigs et al., 'Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements', *Nature* 446 (2007): 908–911.
[437] P. J. Eslinger and A. R. Damasio, 'Severe Disturbance of Higher Cognition after Bilateral Frontal Lobe Ablation: Patient EVE', *Neurology* 35 (1985): 1731–1741.
[438] M. D. Hauser, *Moral Minds*, 247.
[439] Haushofer J, Fehr E (2008). You shouldn't have: your brain on others' crimes. Neuron 60, 738-740.
[440] Koenigs et al., 'Damage', 909.

What would an adequate theory of Moral PFC involve? Wood and Grafman suggest five criteria**:**

> The first of our five proposed criteria is that a theory must be explicit about the information that is stored in the PFC. Does it store information akin to a memory function (representational approach)? Does it store algorithms or computational procedures only for manipulating information stored elsewhere in the brain (processing approach)? Does it do a combination of these things (hybrid approach)? Second, the theory must be consistent with our knowledge of stimulus representation in the brain. If it is not, then the authors must have explained the inconsistency and provided evidence to support its validity. Third, it must be reasonable from an evolutionary perspective. Fourth, it must make predictions that enable verification and invalidation of the model. Fifth, it must be supported by the available physiological data—neuroimaging, electrophysiology, and animal and human lesion research.[441]

The distinction between *processing* and *representational* viewpoints of PFC operation is essential to understanding how the brain organises and facilitates responses to stimuli. It is also essential in terms of an explanation of how the brain makes ethical decisions. Joshua Greene and others (e.g. Patricia Churchland, Paul Churchland and William Casebeer), argue for a processing view (with a control function between rational and emotional decision making), while Jorge Moll and others favour a representational approach (where social knowledge is bound to emotions).[442]

I now consider a number of models for PFC and anterior PFC operation utilising these viewpoints. I do this by first considering these theories in a general sense (so not necessarily related to moral decision making); then by looking at PFC operation(s) that emerge in relation to moral testing.

So what exactly do we mean by a representation structure as opposed to a processing structure?

---

[441] Wood and Grafman, 'Human Prefrontal Cortex', 139.
[442] J. Moll et al., 'The Cognitive Neuroscience of Moral Emotions', in Sinnott-Armstrong, *Moral Psychology*, 3:7.

> We define representations as memories that are localized in neural networks that encode information and, when activated, enable access to this stored information. Processes, on the other hand, are computational procedures or algorithms that are localized in neural networks and are independent of the nature or modality of the stimulus that is being processed.[443]

Representations, then, are activated by certain incoming stimuli with a certain pattern or strength that triggers the knowledge and brain state/function. Those of a process viewpoint might ask how their view is accommodated by a representational structure? Wood and Grafman reply by arguing that '"processes" in cognition are a set of representations that, when activated, remain activated over a period of time—a possibility that is supported by data showing sustained firing by PFC neurons'. [444] On this view, 'process' is a representational construct.

I turn now to a popular processing theory. The *Adaptive Coding Model* is a processing approach that illustrates the role of the PFC in selecting and integrating sensory information (stimuli). [445] This selection and integration utilises the adaptability of the neuron; and so we find fMRI modelling of the same neurons performing different functions.

> In this model, working memory, selective attention and control are simply three different perspectives on the same underlying processing function. The central idea is that, throughout much of prefrontal cortex—certainly including much of the lateral surface—the response properties of single neurons are highly adaptable. Any given cell has the potential to be driven by many different kinds of input—perhaps through the dense interconnections that exist within the prefrontal cortex. In a particular task context, many cells become tuned to code information that is specifically relevant to this task. In this sense, the prefrontal cortex acts as a

[443] Wood and Grafman, 'Human Prefrontal Cortex', 139.
[444] Ibid., 140.
[445] Ibid., 141.

> global workspace or working memory onto which can be written those facts that are needed in a current mental program.[446]

An alternative processing model that does not focus on the adaptability of neurons is the *Attentional Control Model*, which offers an even more mechanistic process account.[447] This model adopts an integrated valuation and execution component:

> There are two mechanisms that monitor behaviour. The *contention scheduler* results in automatic priming of stored knowledge and the *supervisory attentional system* controls the setting of priorities for action. The SAS reflects conscious awareness rather than simple responses to stimuli. The SAS is localized in the PFC; however, the localization of the contention scheduler is unspecified. The SAS can override the contention scheduler when necessary—for example, the ring of a telephone will cause priming of 'answer the phone' behaviour by the contention scheduler, but it might be appropriate for the SAS to override this if the telephone belongs to someone else.[448]

This view is consistent with the biological structure of the PFC, and experiments on behavioural control enhance its plausibility. Interestingly, it uses 'awareness' as the control setting, meaning conscious consideration of incoming stimuli (taking the associative learning performed by the brain and integrating it with consciousness).

Perhaps a more promising alternative / abstract model is the *Connectionist Model*, which is a multi-functional, hierarchical process explanation, with the PFC playing a role at higher functional levels.

> The model considers four levels of the cortical system — cell, module, tissue and global—that integrate learning experiences to produce a coherent functional system. The levels have different functions: the cellular level processes information and modifies neuronal behaviour; the modular level enables computation and learning within a cortical column; the tissue level activates

[446] J. Duncan, 'An Adaptive Coding Model of Neural Function in Prefrontal Cortex', *Nature Reviews Neuroscience* 2, no. 11 (2001): 824.
[447] Ibid.
[448] Wood and Grafman, 141.

> different inputs in parallel and integrates successive learning experiences; and the global level integrates functions from different cortical regions to produce behaviour. [449]

This view (a kind of supervenience-in-action) proposes that units in the PFC correspond to specific sensory or motor events, and are selective for event sequence. [450]

Having briefly covered the basic processing theories of PFC operation, we now consider representation views, beginning with the *guided activation theory* (GAT). GAT supposes a dual level process of representations between the PFC and the posterior cortex in the context of an hierarchical view of the LPFC bias information flow through various operating networks. The theory was developed by Miller and Cohen:

> GAT proposes that the PFC stores representations of task-specific rules, attentional templates and goals. Essentially, the PFC 'directs' activation to bias the activation of goal related representations that are stored in posterior cortex.[451]

In 2013 the theory was upgraded by Cole et al.'s[452] as *flexible hub theory*, which argues that such representations are made possible by flexible hubs in the fronto-parietal brain network (brain regions that rapidly update their pattern of global functional connectivity according to task demands).

The *Temporal Organization Model* is similar to the Connectionist Model in that it has a sequential structure, represented at different levels of a cortical hierarchy, culminating in the most senior directing level, the PFC, which plays a functional role through temporal integration (allowing for continued analysis despite variance in environmental stimuli over time):[453]

---

[449] Ibid., 142.
[450] Ibid., 142.
[451] Ibid., 144.
[452] M. W. Cole et al., 'Multi-Task Connectivity Reveals Flexible Hubs for Adaptive Task Control', *Nature Neuroscience* 16, no. 9 (2013): 1348–55
[453] J. M. Fuster, 'Upper Processing Stages of the Perception: Action Cycle', *Trends in Cognitive Science* 8 (2004): 143–145.

> Sequential behaviour arises from an elaborated perception–action cycle, within which a mediating role is played by internal representations capable of maintaining information about temporal or task context.[454]

This model integrates representational and processing viewpoints:

> Given the emphasis on attention, short term memory and inhibitory control, the model seems to be a processing viewpoint. However, Fuster also describes PFC function in terms of 'motor memory' (schemas), with a hierarchy of motor representations within the PFC. Attention and working memory are properties of the representations (neural networks), rather than explicit 'processes' in terms of computational procedures.[455]

A further alternative is the *Working Memory Model*, created by Patrica Goldman-Rakic. This model suggests that the PFC serves as a working memory structure that keeps stimulus representations active for short periods of time.

> The PFC is part of an integrated network of regions—temporal, parietal, premotor and limbic—that is involved in the representation of stimuli in their absence; this enables behaviour to be guided by internal representations rather than relying on the presence of external stimuli.[456]

The fact that all these theories of PFC operation are supported by various fMRI studies and other scanning and experimental or analytical studies indicates that as yet we really do not know how the PFC coordinates task and makes decisions. What has emerged is agreement on the kinds of tasks it performs – it integrates sensory and memory information towards actions and behaviours utilising information from other areas of the brain.

This quick run through of theories paints a general picture: the PFC is primarily involved in the coordination of decision making dependant on the quality of coded representations determined through other brain areas. What a representation model does best is in allowing

[454] M. M. Botvinick, 'Multilevel Structure in Behaviour and in the Brain: A Model of Fuster's Hierarchy', *Philosophical Transactions of the Royal Society, Biological Sciences* 362 (2009): 1620.
[455] Fuster, 'Upper Processing Stages', 144.
[456] Ibid., 145.

for a linking between cultural norms and specific emotions,[457] with violations of these norms eliciting such emotions. However, such binding, working in much the same way as associative learning, must have a processing construct, and so it is quite likely that the truth here will involve an integrated modular theory.

With these representational and processing theories in mind, we can now explore the PFC in terms of specific theories with moral consequences; and we can do so because of the consistency in the kinds of tasks the PFC performs, regardless of its ultimate organisational structure.

9.3 Somatic Marker Hypothesis

Under the SMH, selected emotion guide cognition via an integrated system of automatic marker signals that relate to body-state structure and regulation. In this theory decision making is not a product of cognition alone. Damasio (1994) constructed this hypothesis on the basis of work with patients displaying a specific type of emotional degradation from focal damage (VMPFC lesions) to their pre-frontal lobe. The VMPFC plays a significant role facilitating information for decision-making.[458]

Perhaps the most interesting element of SMH concerns the utilisation of emotions; something especially relevant to concerns with moral decision-making. Damasio defines an emotion as an ensemble of changes in the body and brain states. These are determined by a specific system that automatically responds to specific external contents of one's perceptions, be they actual, or recalled from memory.[459] This hypothesis extends not only to emotional decision making, but also to rational decision making.[460]

Damasio divides emotions into primary and secondary emotions, based on his experimentation of people with damage to their VMPFC. The primary emotions include happiness, sadness, fear, anger, surprise and disgust; while the secondary emotions are

[457] P. R. Erichlich, *Human Natures: Genes, Cultures, and the Human Prospect* (Washington, DC: Island Press, 2000).
[458] F. Grabenhorst and E. T. Rolls, 'Value, Pleasure and Choice in the Ventral Pre-Frontal Cortex', *Trends in Cognitive Science* 15 (2011): 56–67.
[459] A. Bechara and A. R. Damasio, 'The Somatic Marker Hypothesis: A Neural Theory of Economic Decision', *Games and Economic Behaviour* 52 (2005): 340.
[460] R. Firat and C. M. McPherson, 'Toward an Integrated Science of Morality', in *Handbook of the Sociology of Morality*, ed. S. Hitlin and S. Vaisey (New York: Springer, 2010), 365

more complex social emotions (e.g. pride, jealousy, envy, contempt). Both types are involuntary. Damasio's claim is that that frontal lobe damage tends to affect the secondary emotions.

Different emotions evoke different bodily responses and utilise different brain systems. It is common to have them loosely tied together as elements or components of the limbic system, though recent work has challenged the idea that the limbic system is solely/largely responsible for emotional driven. Domasio's idea is that these emotions are located in the VMPFC:

> Damasio and colleagues observed that patients with ventromedial PFC damage can detect the implications of a social situation, but cannot make appropriate decisions in real life. They suggested that such patients would be unable to mark those implications with a signal that automatically distinguishes advantageous from pernicious actions. The somatic marker model explains why patients with ventromedial PFC damage can still reason about social problems, provided the premises are cast verbally, but fail in natural settings.[461]

Although Damasio's theory has widespread ramifications, many have criticised it as it does not explain more recent findings concerning moral cognition:

> This framework is compatible with contextual effects (although these are not explicitly addressed), integrates cognition and emotion, makes testable predictions, and has been supported by neurophysiological and clinical data. However, it does not explicitly address the role of different PFC sub regions in moral cognition. The relationships between somatic markers and other cortical and limbic regions that have previously been linked to moral cognition are also obscure. Recent evidence from both patients with PFC lesions and healthy individuals has challenged the role of somatic markers in guiding decision making and social behaviours.[462]

[461] Moll et al., 'Neural Basis', 801.
[462] Ibid., 802.

In the light of this challenge Koenigs et al. distinguish brain activity in response to particular types of moral dilemmas**,** offering an enriched account of the PFC and its role in moral decision-making. Their study demonstrated that patients with damaged VPFC could respond normally if the moral task was to recognise and condemn the harmful actions of others; however, they would themselves favour performing just these actions (e.g. pushing a fat man off a bridge to save others).[463] The former recognition of harmful action can be understood as a kind of norm awareness derived from previous experience, stored elsewhere in the brain, and still influential. Similarly, Haidt's social intuitionist model holds that while moral intuitions may engage somatic markers, they can be overridden by conscious post-hoc reasoning.[464]

9.4 Social Response Reversal

Social Response Theory proposes that a neurocognitive system is activated by another's expressions, which stops certain behaviours and reverses the current response in favour of an alternative.[465] If, for example, I see another person is being hurt by my comments then the system kicks in and I (minimally) desist.

James Blair, from the Institute of Cognitive Neuroscience in London, developed SRT while studying aggression.[466] Blair's evidence suggests a hierarchical relation between cognition and emotion, whereby specific models suppress aggressive behaviour. It follows that if these models are damaged, various types of aggressive behaviour result. Blair demonstrated these models operate regardless of the nature of the produced social action:

> Humans have a considerable facility to adapt their behaviour in a manner that is appropriate to social or societal context. A failure of this ability can lead to social exclusion and is a feature of disorders such as psychopathy and disruptive behaviour disorder. We investigated the neural basis of this ability using a customized video game played by 12 healthy participants in an fMRI scanner. Two conditions involved extreme examples of context-appropriate action:

[463] Koenigs et al., 'Damage ', 910.
[464] Firat and McPherson, 'Toward an Integrated Science', 365.
[465] Ibid., 1136.
[466] James R. Blair, 'Neurocognitive Models of Aggression, the Antisocial Personality Disorders, and Psychopathy', *Journal of Neurology, Neurosurgery and Psychiatry* 71 (2001): 727.

> shooting an aggressive humanoid assailant or healing a passive wounded person. Two control conditions involved carefully matched stimuli paired with inappropriate actions: shooting the person or healing the assailant. Surprisingly, the same circuit, including the amygdala and ventromedial prefrontal cortex, was activated when participants acted in a context-appropriate manner, whether being compassionate towards an injured conspecific or aggressive towards a violent assailant. The findings indicate a common system that guides behavioural expression appropriate to social or societal context irrespective of its aggressive or compassionate nature.[467]

For Blair, psychopathy arises from the fact (however caused) that the system that recognises distress in others fails to function correctly**.** This failure typically results from damage to the VMPFC.[468]

A later study by Koenigs et al. has given further support for the role of the VMPFC determining emotional content in moral decision making; while a more recent study by Tabber-Thomas et al. in 2014 showed the same results with patients who suffered VMPFC damage at an early age.[469]

9.5 The SEC Framework

> *Our structured event complex (SEC) framework proposes that the PFC stores unique forms of knowledge. A SEC is a goal-oriented set of events that is structured in sequence and represents thematic knowledge, morals, abstractions, concepts, social rules, event features, event boundaries and grammars. The stored characteristics of these representations form the bases for the strength of representation in memory and the relationships between SEC representations.* [470]

Wood and Grafman's SEC framework holds that the PFC is more than a facilitator or conductor of knowledge; it is a storage device that houses representations of the external

[467] J. A. King et al., 'Doing the Right Thing: A Common Neural Circuit for Appropriate Violent or Compassionate Behaviour', *Neuroimage* 30 (2006): 1069.
[468] A recent study has shown that it may also result from a lack of learning[468] (i.e. experience playing with others as a child), though the VMPFC is not involved in such learning. (F. Cushman, 'The Neural Basis of Morality: Not Just Where, But When', *Brain,* 137 (2014): 974.)
[469] Interestingly, the early age patients did not condemn self-interested actions. This was likely due to a) a lack of knowledge of norms elder patients had acquired, or b) the VMPFC damage to younger people shows it does actually play a deeper role than mere facilitation.
[470] Wood and Grafman, 'Human Prefrontal Cortex', 143.

world. SEC representations are basically long-term memories of event sequences that guide the perception and execution of goal-oriented activities.[471] On this view, the relationship between cognition and emotion is hierarchical because the representations control emotional responses. Later research in this area suggests *counterfactual* representations, so enabling theorising or imagining outcomes to be included.[472] This framework makes specific predictions regarding the properties and locations of SECs in the PFC, and it has fMRI support:

> An SEC representation includes situational knowledge abstracted across events (going to a concert) and the temporal organization of events (making a reservation, dressing up, and so on). Activated SECs sequentially bind representations of objects, actions and spatial maps stored in posterior brain regions. The SEC framework predicts that different subdivisions of the PFC store different types of content or domains of event knowledge. Clinical and neuroimaging evidence supports this prediction, showing that different PFC regions are involved in representing social and emotional SECs (ventromedial PFC), novel or multi-tasking event sequences (anterior PFC) or overlearned sequences (more posterior PFC regions).[473]

An event is, together, any activity that begins a causal sequence, the activities involved in achieving that sequence, and a final activity, signifying the end of the event.[474] The more often the sequence is performed, the more automatic the behaviour becomes. The idea is that such developed SECs define the possibilities for thinking of alternative possibilities.[475]

Moll et al. are sceptical about this capacity of this framework when it comes to accounting for moral values and moral decision making:

> Although this framework has clear implications for moral cognition, these rely on the hypothesis that the PFC stores the situational and temporal context of social

[471] Moll et al., 'Neural Basis', 803.
[472] A. K. Barbey, F. Krueger, and J. Grafman, 'Structured Event Complexes in the Medial Prefrontal Cortex Support Counterfactual Representations for Future Planning', *Philosophical Transactions of the Royal Society, Biological Sciences* 362 (2009): 1291.
[473] Moll et al., 'Neural Basis', 803.
[474] Barbey, Krueger, and Grafman, 'Structured Event Complexes', 1292.
[475] Ibid., 1295.

> knowledge. The SEC framework does not predict how PFC regions interact with limbic areas and other cortical regions to give rise to a range of moral cognitive phenomena, such as moral values and moral emotions.[476]

9.6 The EFEC Framework

The *event feature emotion complex* (EFEC) framework

> … postulates that moral cognitive and behavioural phenomena arise from the binding of three main components: structured event knowledge (provided by context-dependent representations in prefrontal subregions), social perceptual and functional features (stored in the posterior and anterior sectors of the temporal cortex) and central motive or basic emotional states (such as aggressiveness, sadness, attachment or sexual arousal, represented in limbic and paralimbic regions)…These component representations give rise to a 'gestalt' experience by way of temporal synchronization.[477]

On this view, different neural groups hold knowledge and motivation states; which knowledge and states, given external stimuli, give rise to a particular moral output. Here there is no post-hoc reasoning, but rather an integrated processing mechanism, which facilitates content specific process reasoning (i.e., long-term interests and cultural factors). The theory suggests that moral decision making spreads to usage of brain areas beyond the PFC:

> The evidence discussed above strongly indicates that the neural mechanisms of moral cognition are not restricted to the PFC, limbic areas or any other brain region. We propose a new representational neural architecture, designed to circumvent the limitations of previous frameworks. In our view, moral cognitive phenomena emerge from the integration of content and context dependent representations in cortical–limbic networks.[478]

[476] Moll et al., 'Neural Basis', 802.
[477] Ibid., 805.
[478] Ibid.

9.7 Dual Process and Modular Myopia

Joshua Greene, a philosopher and neuropsychologist, endorses and exploits reductionism when it comes to uncovering and understanding the causal origin of moral cognition and decision making. His particular technique is to utilise moral dilemmas from traditional moral philosophy (i.e. the Trolley Problem and the Footbridge Dilemma) [479] and place them under the scrutiny of fMRI studies:

> Just as visual illusions reveal the structure of visual cognition, bizarre moral dilemmas reveal the structure of moral cognition. They are moral illusions, revealing for the manner in which they mislead us.[480]

'Dual process' refers to the idea that both cognitive/reasoning and emotional processes can determine moral judgment, and that which process is selected depends on the environmental cues. Greene refers to the emotional processes as 'automatic mode', and to the cognitive/reasoning processes as 'manual mode', and insists that these are real, not merely abstract, modalities.

> Manual mode is not an abstract thing. It is a set of neural networks, based primarily in the prefrontal cortex, that enables humans to engage in conscious and controlled reasoning and planning… It is a cost benefit reasoning system that aims for optimal consequences.[481]

In one version of the trolley problem a person is asked to flick a switch that would see a train kill one person as opposed to killing five people. Testing established that most people are likely to opt to save the five persons, suggesting a utilitarian response. However, in another version known as the footbridge dilemma, people were asked to push a fat man off a bridge in order to save five people below from being hit by a train. Greene found that the act of having to push another human off a bridge – a very personal physical engagement – reversed the results. Greene's explanation is that the emotional systems of the brain are activated by the potentiality of the personal engagement, leading to a non-utilitarian moral

[479] P. Foot, 'The Problem of Abortion and the Doctrine of Double Effect', *Oxford Review* 5 (1967): 5–15; Thomson, 'Trolley Problem'.
[480] Greene, *Moral Tribes*, 252.
[481] Greene, *Moral Tribes*, 202.

decision. He concludes that people exhibit a characteristically consequentialist response to the trolley problem (flick a switch), and a characteristically deontological response to the footbridge case. This is precisely what the neural mapping of emotional and cognitive areas showed in his 2004 fMRI study. The emotional areas included the posterior cingulate cortex (PCC), the medial PFC, and the amygdala. The cognitive areas included the DLPFC, and the inferior parietal lobe.[482]

On the basis of these findings Greene claims that the origins and logic of traditional moral theories – utilitarian, deontological, and Aristotelian – have their origins in the neuroscience of the brain:

> We can use manual mode setting to specifically describe out automatic settings (Aristotle) We can use manual mode thinking to justify our automatic settings (Kant). We can use manual mode thinking to transcend the limitations of our automatic settings (Bentham and Mill).[483]

Greene's central finding from his 2004 fMRI study is that utilitarian moral judgment is associated with increased activity in the DLPFC, the brain region most closely associated with cognitive control. This result was specifically predicted by the DPT, connecting literature on the neuroscience of cognitive control (Miller and Cohen, 2001), with research on moral judgment:

> The dual-process theory claims that dilemmas like the *footbridge* case elicit a potent negative emotional response to the action that supports the characteristically deontological response (It's wrong, even though it will save more lives) as well as a competing controlled cognitive response that supports the characteristically utilitarian conclusion (It's morally acceptable because it will save more lives)… If the dual-process theory is correct, then deontology is fundamentally an intuitive (automatic) philosophy… Consequentialism, in contrast, is *geek morality*. It's what you get when you turn the problem of moral thinking over to the brain's manual mode… Utilitarian judgements depend more

[482]Greene, 'Secret Joke', 44.
[483] Greene, *Moral Tribes*, 331.

> on cognitive control. Removing time pressure and encouraging deliberation increases utilitarian judgement.[484]

These results led Greene to ask the question: *can reflecting on a moral question change one's mind*? A study by Suter and Hertwig claimed to show that decreased deliberation decreases the likelihood of utilitarian judgment-making,[485] and Greene wished to test this finding in light of DPT. Paxton, Ungar and Greene, found the answer to this question was *yes*. Through two experiments (cognitive reflection tests), they demonstrated that reflection, which involved intuitive and moral reasoning, as opposed to automatic emotional influence, increased the likelihood of utilitarian moral reasoning and decision-making. The key determinates of reasoned reflection were sensitivity to argument strength, and extended temporal duration.[486]

Greene also posits the *modular myopia theory* (MMT), which seeks to explain (and fill) some of the gaps in DPT. The idea is that it accounts for why emotional systems can, in some cases, overrule the, cognitive, manual mode:

> First, our brains have a cognitive subsystem, a 'module,' that monitors our behavioural plans and sounds an emotional alarm bell when we contemplate harming other people**.** Second, this alarm system is myopic, because it is blind to harmful side effects… it's an action plan inspection system, a device for keeping us from being casually violent… What we are hypothesizing is a little alarm system, an automatic setting that provides a check on the potentially dangerous plans drawn up by the outcome maximizing manual mode… Aligned with the dual process brain we can explain why we care less about harms caused as side effects, than as a means to an end… The side effects don't set off the alarm because the harmful event is not on the primary chain of the action plan.[487]

Multiple studies align with the core of Greene's findings: factors like spatial proximity, argument strength and temporal duration, direct the selection of which part of the brain it is

[484] Ibid., 127.
[485] R. S. Suter and R. Hertwig, 'Time and Moral Judgment', *Cognition* 119 (2011): 454–458.
[486] J. M. Paxton, L. Ungar, and J. D. Greene, 'Reflection and Reasoning in Moral Judgment', *Cognitive Science* 36 (2011): 164.
[487] Greene, *Moral Tribes*, 247.

that ends up making a moral decision. These four studies come from 'trolleyology' experimentation:[488]

- People who are more empathetic, or induced to be more empathetic, give more deontological responses.[489]
- Inducing mirth (the positive emotion associated with humour, here thought to counter-act negative emotional responses) increases consequentialist responses.[490]
- Performing a distracting secondary task (i.e. being under cognitive load) reduces consequentialist responses.[491]
- Individuals who generally favour effortful thinking over intuitive thinking are more likely to give consequentialist responses.[492]

Greene is not blind to the potential practicality of his work. He urges us to use this knowledge towards what he calls a 'common currency'. As it becomes widely known how we go about thinking morally, there can emerge a deeper commonality that allows us to transcend moral localism of the kind (in a way, paradoxically) deontology engenders.

> To take my favorite example (Peter Singer's, too), I think that deontological thinking is a major obstacle to ending poverty (Greene, 2003). Nature didn't design us to behave morally in a world in which one can save the life of a distant stranger at very little cost. If Singer (2005) and I are correct, we humans rely too much on our emotional intuitions to tell us when we have a moral obligation to help and when we don't. We feel an obligation to help when a victim of misfortune is right in front of us (drowning baby), but not when the victim is a distant, 'statistical' one (Small & Loewenstien, 2003; Slovic, 2007) on the other side of the world (giving to Oxfam).

[488] For a more extensive summary of recent findings, see Greene, 'Beyond Point-and-Shoot Morality', 714.
[489] P. Conway and B. Garwronski, 'Deontological and Utilitarian Inclinations In Moral Decision Making: A Process Dissociation Approach', *Journal of Personality and Social Psychology* 104, no. 2 (2013): 216–235.
[490] N. Strohminger et al., 'Divergent Effects of Different Positive Emotions on Moral Judgment', *Cognition* 119, no. 2 (2011): 295–300.
[491] Conway and Garwronski, 'Deontological and Utilitarian Inclinations', 296.
[492] D. M. Bartels, 'Principled Moral Sentiment and the Flexibility of Moral and Decision Making', *Cognition* 108 (2008): 381–417.

The practicality here is obvious, and not just for poverty alleviation, but for an Anthropocene environmental ethics. We can now explain in a fully naturalistic way why we are so poor at caring for the wider biosphere—our brains simply do not think this way unless such conditions present themselves to us 'face on':

> In sum, the dual process theory of moral judgement, which emphasizes both emotional intuition and controlled cognition, is supported by multiple fMRI studies using different behavioural paradigms, multiple behavioural studies of neurological patients, and a variety of behavioural studies using both experimental manipulations and individual differences.[493]

In my view, Greene makes a powerful case for DPT**.** Furthermore, the generality of DPT can easily be aligned with other studies – after all, it is clearly true that emotion often gets in the way of clear headed reason, and equally clear that localist concerns inform much moral cognition and decision-making.[494] The trouble is that the further one moves from psychology and enters directly into the realm of neuroscience, the more ambiguous specific causal origins for certain behaviours becomes. Phelps, Lempert and Sokol-Hessner, dispute the value of categorising the brain into a dual system:

> What is emerging is clearly incompatible with the notion of two systems. Rather the literature suggests that there are multiple neural circuits underlying the modulation of decision making by emotion of affect…the specific neural circuits involved vary depending on which affective component is engaged and which decision variables are assessed.[495]

While I can see the point of this scepticism, these finer details might not be required to draw meaningful conclusions. DPT may be descriptively adequate when it comes to what goes on in the brain and the ultimate behaviours produced, in much the same way Kahneman and Trevsky's psychological experimentation produces *prospect theory* (something I take up in the following chapter).

---

[493] Greene, 'Cognitive Neuroscience'.
[494] Edward B. Royzman and J. Baron, 'The Preference for Indirect Harm', *Social Justice Research* 15 (2002): 165–-84.
[495] Phelps, Lempert, and Sokol-Hessner, 'Emotion and Decision Making', 281.

The literature in this chapter has revealed obvious, but also more subtle contradictions. Each study simply read on its own tends to reveal a 'eureka insight', but clearly when taken together such causality is weakened by seemingly valid opposing theories. The representational vs. processing understanding of the moral PFC was a prime example in this respect, and indeed it ultimately weighs upon those other frameworks explored. What these theories have to do with one another then is that they each offer a particular kind of explanation being the naturalistic processes occurring within the PFC (and other relevant areas of the brain) when we perform certain kinds of moral decision making, but as particulate as each is we can see that when taken together they are vague with respect to all the brain does in moral decision making, and so we cannot really appreciate this vagueness unless we discern each theory and then simply state the next one after it. However, such vagueness shouldn't be interpreted as dismissive of these studies. Rather, in those specifics where mechanistic explanations were revealed we are now able to progress these findings when it comes to the application of environmental ethics, through asking a newfound relevant question in each respect. Before we can do this though, we need to consider methodologies through which we can bridge such empirical theories to environmental issues.

Chapter 10: Environmental Decision Making and Policy Construction

10.1 Beyond Traditional Methods: CBA and PP

> *Not only do values and their material basis constrain social problem-solving, they are at the same time facts that give substance to the processes of figuring out what to do—facts such as that our children matter to us. And that we care about their well-being; that we care about our clan. Relative to these values, some solutions to social problems are better than others, as a matter of fact; relative to these values, practical policy decisions can be negotiated.*[496]

I have argued that the relevance of neuroscience to morality does not depend on rejecting the is-ought distinction. Any moral ought, whatever its factual origins, is normative in the traditional philosophical sense. Even so, I have argued that such oughts, in terms of their real applicability or social operationalizability, still answer to realities determined by the mechanics of behavioural causation. And I have further suggested that reflective awareness of these realities potentially enables us to better and more effectively tailor our policy-making.

One example of this general orientation can be drawn from the work of Eleanor Ostrom, who in 2009 was awarded the Nobel Prize in economics for showing that people, if their institutional organisation possesses certain design features, are capable of managing common property resources,[497] these 'design features' are products of social evolutionary history as held in our brain's neural functionality.

The opportunity is to identify consistent patterns of causation that result in certain types of decisions being made – which, for this thesis, means those patterns of causation that inform

[496] Churchland, *Brainstrust*, 9.
[497] Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, 1990.

our environmental decision-making – with the ambition of delivering up an implementable and adequate environmental policy framework for dealing with the sustainability in the Anthropocene.[498]

When it comes to decision-making there are currently a number of tools used to assist in making more informed, objective decisions, many from the field of economics. As we will see, these are inadequate insofar as they fail to incorporate underlying bio-causal factors of the kind explored here. Computational accounts must consider the underlying objective causality.

Cost benefit analysis (CBA) is perhaps the best known formal decision-making tool in use today, despite the fact that it has received much criticism for its inability to incorporate social and environmental ethics into its primarily economic construct.[499] The economists' defence to these challenges is typically simply to invoke the 'invisible hand'. In response many have appealed to the *precautionary principle* (PP), and especially when it comes to environmental policymaking. With this formulation seemingly too close to 'radical environmentalism' for many economists, their response (those who think one necessary)[500] has been to develop the idea of 'option value', which is 'the value of maintaining flexibility for future decisions and avoiding irreversible, or costly to reverse, outcomes'.[501] Option value or the PP can be applied to ecosystem management:

> There is an option value for conserving an ecosystem because current development will foreclose the option for future conservation, whereas current conservation does not foreclose the option of future development. Option value can make it desirable to conserve an ecosystem even though the expected value of development exceeds the expected value of conservation. [502]

Even with this addition, CBA remains the tool of those primarily concerned with economic gain, and (because of the so-called 'discount rate') generally immediate economic gain or

[498] For a succinct explanation of potentially relevant environmental ethics, see Table A-1.
[499] Such attempts as there are - as with contingent valuation surveys – depend on question-begging assumptions and – typically – in simply ignoring the fact that very many people refuse to participate in such surveying in the first place.
[500] Aldred, 'Justifying Precautionary Policies', 132.
[501] Polasky et al., 'Decision-Making under Great Uncertainty', 400.
[502] Ibid.

electoral advantage (politics), suggesting that it is not so much the tool that matters, as the motivations of the people who utilise it.

The advantages and disadvantages of PP and CBA when it comes to guiding policy, and environmental policy in particular, have already been well established,[503] and will not be explored here**.** Instead we will take the opportunity to uncover and explore those 'design features' that shape the cognition and decision-making of those who act in environmentally harmful ways. Such design features are legitimately conceivable in a naturalistic and reductionist framing. While *decision theory* (DT) in the form of *expected utility theory* (EUT) has been used in classical economics and environmental decision making, its full usefulness requires understanding our cognition.

This has been recognised. There have been psychological studies that reveal real agents deviations from the demands of objective mathematical rationality. Kahneman and Tversky have explored how people actually make decisions, and found deviations from what EUT suggests. Their modified *prospect theory* is testament to the potential of descriptive accounts of the kind this thesis endorses. Nicholas C. Barberis, in an excellent review of the 30 years since the birth of prospect theory, outlines the elements of this understanding:[504]

- Reference dependence: People derive utility from gains and losses, measured relative to some reference point, rather than from absolute levels of wealth (i.e.: Pope and Schweitzer (2011) show that prospect theory plays a role even in the behaviour of highly experienced and well-incentivised professionals: in particular, professional golfers are significantly more likely to make a putt for par than a putt for scores other than par, a finding that is consistent with loss aversion relative to the reference point of par.)
- Loss aversion: The idea that people are much more sensitive to losses—even small losses—than to gains of the same magnitude.

[503] Aldred, 'Justifying Precautionary Policies', 137.
[504] Nicholas C. Barberis, 'Thirty Years of Prospect Theory in Economics: A Review and Assessment', *Journal of Economic Perspectives* 27, no. 1 (2013): 175–189.

- Diminishing sensitivity: Implies that while replacing a $100 gain (or loss) with a $200 gain (or loss) has a significant utility impact, replacing a $1,000 gain (or loss) with a $1,100 gain (or loss) has a smaller impact.
- Probability weighting: People do not weight outcomes by their objective probabilities but rather by transformed probabilities or 'decision weights'. The decision weights are computed with the help of a 'weighting function' whose argument is an objective probability.

My strategy here is not to question DT on grounds of empirical psychology (as does prospect theory), but to exploit the causality of the biology that realises those decisions**.** I argue for the legitimacy and value-adding of an *internalist* (sociobiological reductionist) approach. I argue this kind of explanation may allow for improvements in the underlying theory of decision making because it is no longer a case of 'this is what we actually do' *v.* probability (Kahneman and Tversky), but rather 'this is why we do it', (e.g. neuroscience/module X, cued on environment Y). Our research should not just be informed by test case scenarios that confirm human bias in decision making, it should be informed by underlying causal hypotheses based on what we know about the origins of our decision making, given our evolved neural structures. If we know 'why we do it' we can better estimate what decision-making changes are possible. Section 10.5 is directed to this end. To get there I look first at standard decision analysis and its treatment of ethics.

10.2 Ethics in the Context of Decision Theory

Let us briefly recap the original applicative approach explored in Robert Traer's 'doing environmental ethics':[505]

1. Construct an ethical presumption
2. Consider the consequences
3. Make your decision

The 'ethical presumption' in Step 1 is simply an ethical claim (i.e. that X is good/bad/just/unjust, or that X has an intrinsic value). The 'consequences' in Step 2 are an

[505] Traer, *Doing Environmental Ethics*, 163.

ethical consideration of the future outcomes should the presumption be applied to reality, while 'Make your decision' (Step 3) involves weighing up 2 and 3 to reach a conclusion. The importance of these considerations is without question. What I want to add to it is a deeper consideration of those factors which determine the strength or sense of the move from 2 to 3; which is to say, those factors which inform, shape and constrain the applicability or operationizability of the final decision.[506]

Standard DT incorporates a probability of success for an action into action determination, and in this way manages risk:

> For each act–state pair (or outcome) we assign a probability and a utility, then we multiply these together and sum the products across each act. The resulting sum we call the expected utility of the act. Standard decision theory then tells us to choose the act with the greatest expected utility (if there is such an act).[507]

DT has significant advantages. First**,** it provides a clear statement of the problem and objective; second, it incorporates quantitative and economic information under a normative guide; and third, it offers a transparent and repeatable process in a given situation. [508]

A major problem for the theory is that the information required to deduce such probabilities is often unavailable, but for our purposes an even greater problem concerns its treatment of ethics. The use of formal techniques must not negate or replace rational deliberation in the sense described by Truer. Ethics comes before applicative technique; otherwise DT may give us unethical results, especially under its standard form. For example, if we had limited resources and DT told us to preserve a struggling species and disregard another species that would go extinct if resources were allocated in this way, the ethics of this trade-off (to let a species go extinct) should decide if this was acceptable, not the results of DT. Mark Colyvan makes this error (favouring strict DT) with his 'backwards' interpretation of the relation between ethics and DT:

---

[506] It could be argued that 2 and 3 could concern this space, but it should already be clear that they do not in the capacity researched here; they refer to developing the ethic and ethical consequences, a normative gambit, whereas the work here concerns descriptive limitations and possibilities.
[507] Mark Colyvan, Damian Cox, and Katie Steele, 'Modelling the Moral Dimension of Decisions', *NOUS* 44, no. 3 (2010): 504.
[508] Polasky et al., 'Decision-Making under Great Uncertainty', 400.

> The relevant ethical theories may be well motivated and may even deliver the right results when resources are not limited and where there is no uncertainty (i.e. where we know what the result of our actions will be). But sadly we do not live in such a world and in this, the actual world, such ethical theories are useless... To think that ethics alone can tell us what to do in an uncertain world is to make a very dangerous mistake: it is to confuse ethics with DT.[509]

As I see it, Colyvan has misplaced the causal relation between ethics and DT, and so breached the is-ought distinction (the 'is' in this case, an objective mathematical equation as I see it).

Colyvan's reliance on DT is flawed in another sense. 'The actual world' refers to descriptive theory; that is, to the work of K&T on how we actually make decisions. However DT of Colyvan's kind is itself normative with its classic understanding of expected utility, and so removed from the actual world of K&T. Colyvan speaks of the actual world from the point of view of (DT), to justify the fact that ethics cannot be put before normative views (DT). This does not make much sense.

The EUT found in DT is not normative, but rather mathematically objective. Why? The inputs in standard DT are normative, a set of beliefs, but the solution from EUT is a product of mathematical determination based on probabilities. Thus the math of EUT/DT '('is') breaches the is-ought distinction, *if* it is used to pick between ethical views.

It is commonly assumed that EUT and *Bayesian analysis* are normative in that they help to define what we ought or should do. But this common belief is incorrect. What these techniques can do for any proposed 'ought' strategy, is to give the probability of its successful application. This is a factual claim.

With this in mind, it is little wonder Sven Ove Hansson challenges the underlying cause-effect assumptions inherent in DT. Hansson holds that the capacity of DT to handle risk assessment is a topic that should be considered a part of moral philosophy.

[509] Mark Colyvan, 'Environmental Philosophy: Beyond Environmental Ethics', *Arts* 29 (2007): 100.

Consider the Trolley problem, which Hansson criticises for ignoring risk-taking, and implying a far too simple deterministic causation.[510] If we were to take on Hansson's caution, without adopting the methodological distinctions put forth pertaining to 'can and cannot', we would actually be rejecting the informative factual information that Greene, amongst many others, has uncovered about the human brain when it comes to making such decisions. Such information has the potential to allow moral theorising to incorporate such aspects as risk taking. Hansson correctly points out that the consequences of a single action do not cover all the productive connections that it has with future events.[511] However, such criticism should not undermine the use of a reductive ontology and epistemology to determine meaningful cause-and-effect relations where possible, and attempts to incorporate them where useful. To appreciate the difference in views here we need to further unpack DT.

DT has three forms: (a) *Normative*; where the focus is on rational choice (i.e. via EUT, which performs clear probability assessments via mathematical formula); (b) *Descriptive*; where the focus is on how real people make judgments and decisions (i.e. Prospect Theory), and (c) *Prescriptive*; where the focus is on using normative models, but with awareness of the limitations of human judgement and the practical problems of implementing a rational model in a complex world.[512]

DT as commonly discussed refers to (a), because it starts with a goal that is normative (in this case, the preservation of the environment in circumstance Y), and then assumes there is a set of alternative actions X1, X2 etc., to meet it. However, the assumption of normative DT is a decision maker who is fully informed, accurate in computation, and rational in that they will follow statistical results over gut intuitions. This is generally unrealistic, as we know from observation, but evens more so given emerging knowledge of how our brains actually operate. Here Tversky and Kahneman's work established two things:

> First it showed that people making decisions relying only on their intuition were subject to many errors that they would recognise upon reflecting on what they had done. This emphasised the need for a formal procedure such as decision

[510] Hansson, 'Harmful Influence', 587.
[511] Ibid., 591.
[512] Edwards, Miles, and Winterfeldt, *Advances in Decision Analysis*, 5.

> analysis to assist in making important decisions. The second contribution was to show the necessity for those who are assisting in the probability and preference assessments to be aware of the many pitfalls that are characteristics of human thought.[513]

Informing probability calculations with our understanding of naturalist facts about brains and how they operate, seems like a great idea. For example, it has been suggested that prospect theory explains why people gamble in casinos for a longer period of time than they were originally intending, particularly when losing.[514] Given this we might hope that once people are made aware of their inherent bias, especially in formal considered decision-making frameworks, they might make better decisions**.**

At this point it seems that (c), a prescriptive view, is the most reasonable, but in fact it too is flawed, precisely because ethics has been left out of the equation. DT in this account is an example of a breach of the is-ought distinction, since 'to determine what rational people *ought* to do, it is not sufficient to present purely factual evidence about how people do *actually* behave'.[515] One is simply relying on mathematical relational proportionality instead of behavioural observation to determine what we *ought* to do.

How one might integrate DT with more detailed versions of ethical theories beyond the utilitarianism that appears implicit within it, is something that has been explored by Colyvan, Cox and Steele, who have tried to model virtue ethics, deontological ethics, and a more complex form of utilitarianism, into a standard DT framework.[516]

> …we need the epistemological and ethical theories to be spelled out in ways that enable them to be accommodated in a decision-theory framework.[517]

While they attempt to model DT so as to incorporate these ethics, what I am exploring are the elements of the biological and cultural structure implicit in decision making itself. The advantage of this approach is that we need not 'spell out' these ethical theories in

[513] Ronald A. Howard, 'The Foundations of Decision Analysis Revisited', in Edwards, Miles, and Winterfeldt, *Advances in Decision Analysis*, 36.
[514] Barberis, 'Thirty Years of Prospect Theory', 180.
[515] Martin Peterson, *An Introduction to Decision Theory* (New York: Cambridge University Press, 2009), 294.
[516] Colyvan, Cox, and Steele, 'Modelling the Moral Dimension', 503.
[517] Ibid., p. 505

mathematical terms, but rather use the explored causal linkages (which may or may not represent these theories in actuality) to assess their likelihood of success in the 'real world'.

At this point we are left wondering if there is a place for DT in the application of ethics? Is there a more robust decision-making framework; one that does not breach the is-ought distinction; and in which DT has a place? I think there is, but to see why we need to look at the idea of *bounded rationality*.

10.3 Bounded Rationality

CBA, PP and DT frameworks deal with uncertainty differently, but in the case of DT uncertainty gives rise to different formats beyond the normative, descriptive, prescriptive classifications. These formats have their grounding in a debate between *evidential* DT (which holds actions ought to be evaluated in terms of the evidence they provide for thinking that desirable outcomes will arrive), and *causal* DT (which holds actions should be determined on the basis their ability to causally promote certain outcomes).[518] James M. Joyce defends the view that any adequate account of rational decision making *must* take a decision maker's beliefs about causal relations into account.[519] Although Joyce has some reservations on the input of descriptive sciences, he maintains that the foundations of both views of DT are found in Bayesianism, and that of real-valued expected utility functions.[520]

On the Bayesian view, decision makers are assumed to rely on subjective probabilistic beliefs, reducing decision problems under uncertainty to decisions under risk, so that a unique probability can be attached to all possible outcomes or states of the world.[521] Others, following Von Neumann and Savage, draw a line between *uncertainty* and *risk*. For them, 'risk' concerns subjective expected utility (SEU).[522]

Thus far I have only referred to the standard EUT/SEU theory and not the Bayesian alternative. It is worth unpacking the Bayesian theorem, before moving into descriptive

[518] James M. Joyce, *The Foundations of Causal Decision Theory* (Cambridge University Press, 1999), 4.
[519] Ibid., 4.
[520] Ibid., 4.
[521] Aldred, 'Justifying Precautionary Policies', 132.
[522] Ibid., 132.

theories of bounded rationality, which arise on the back of such information. This is because the Bayesian analysis, like EUT, allows us to precisely reveal where bounded rationality comes into play, as such examples skew from the mathematically correct decision as specified by EUT or Bayesian analysis.

To begin with, let us distinguish between EUTs and Bayesian approaches. Martin Peterson defines the typology of approaches in Bayesian theory, and then internalist and externalist non-Bayesian views of the EUT kind:

> A forty-year-old woman seeking advice about whether to, say, divorce her husband is likely to get very different answers from the Bayesian and his critics. The Bayesian will advise the woman to first figure out what her preferences are over a very large set of risky acts, including the one she is thinking about performing, and then just make sure that all preferences are consistent with certain structural requirements. Then, as long as none of the structural requirements is violated, the woman is free to do whatever, she likes, no matter what her beliefs and desires actually are. The non-Bayesian externalist would advise her to consider the objective probability that she will be happy if she divorces her husband, whereas the non-Bayesian internalist will advise the woman to first assign numerical utilities and probabilities to her desires and beliefs, and then aggregate them into a decision by apply the principle of maximising expected utility.[523]

As Peterson says, the Bayesian analyst is concerned with three axioms to inform preferential structure in decision making. These axioms are unpacked below by Till Grüne-Yanoff, culminating in the standard Bayesian formula:[524]

- The individual has a coherent set of probabilistic beliefs. Coherence here means compliance with the mathematical laws of probability. These laws are the same as those for objective probability, which are known from the frequencies of events involving mechanical devices like dice and coins.

[523] Peterson, *Introduction,* 210.
[524] Till Grüne-Yanoff, 'Bounded Rationality', *Philosophy Compass* 2/3 (2007): 537.

- The Bayesian subject has a complete set of probabilistic beliefs. In other words, to each proposition he assigns a subjective probability. A Bayesian subject has a (degree of) belief about everything.
- When exposed to new evidence, the Bayesian subject changes his (her) beliefs in accordance with his (her) conditional probabilities. Conditional probabilities are denoted p(|), and p(A|B) is the probability that A, given that B is true.

A massive amount of evidence has been gathered to show that people often violate predictions derived from EUT and Basyesian axioms,[525] and do so because of the *heuristics* they use (See Section 5.4 above). Such heuristics are instances of bounded rationality, and generate systematic judgement errors/biases. There are two common heuristics: the *representativeness* heuristic and the *availability* heuristic. 'Representativeness' refers to the process of using cues to associate a target event or quantity with similar targets.[526] A prominent example is to be found in the CIA pilot case. If a pilot can spot an enemy plane with .8 accuracy, but there are six times as many of one enemy plane Y as opposed to another X, a normal representative of the pilots accuracy would be .8. However, correct Bayesian analysis reveals the chance of correctly identifying a plane X is .44. Further examples of representative errors arise in cases of failure regress, inability to determine correct sample size, and in misconceived ideas of randomness.[527] 'Availability' concerns the brain's ability to access or recall information correctly. It has been shown that we overemphasise certain emotionally relevant cues, and popular information, over dull, less considered causes. [528]

From this understanding we can appreciate the definition of bounded rationality as concerning theories of behaviour that deviate from EUT or Bayesian logic. But if this is all we have to show then this definition is limited, for we can't get inside the causes of such (ir)rationality, and so to what is going on in the brain. For this reason some have taken bounded rationality a little further, attempting to give realistic accounts of the underlying cognitive procedures that result in such behaviours. In this context, the approach may be that of 'external theory' or 'internal theory'. The former involves measuring the objects movement or the number of objects movements etc., while the latter

[525] Ibid., 537.
[526] Edwards, Miles, and Winterfeldt, *Advances in Decision Analysis*, 141.
[527] Ibid., 141.
[528] Ibid., 142.

involves considering the internal causes for why they move at all. Both are forms of objective science, and the question whether can they be brought together.

Peterson discusses internalist and externalist views of EUT, but we can also appreciate Bayesian deviations. Grüne-Yanoff refers to these as nonconventional, non-expected utilities. 'Non-conventional' simply means those theories that do not just show deviations from the mean, but also hypothesise how the actual cognitive procedures might work. The push of this thesis is to insist on including not only externalist theories (which, of course, do yield insights into what might be going on internally), but also potentially direct internalist accounts from neuroscience that causally account for such limitations.

Stepping back from these technical points, consider the kind of natural limitations that arise:[529]

- Limited knowledge of the world;
- Limited ability to evoke this knowledge;
- Limited ability to work out consequences of actions;
- Limited ability to conjure up possible courses of action;
- Limited ability to cope with uncertainty;
- Limited ability to adjudicate among competing wants.

These limitations constitute the subject matter of bounded rationality, and an adequate theory of these will have the shape Grüne-Yanoff outlines:

> Any theory of bounded rationality that deserves its name should provide an account of (i) what the limits of human rationality are; (ii) what effects these limitations have on the process and the outcome of deliberation; and (iii) why these effects are realised in some contexts and not in others.[530]

Meeting these demands raises questions of the extent of model realism. Realism here can be approached pragmatically, as usefully summarised by Milton Friedman:

[529] Herbert A. Simon, 'Bounded Rationality in Social Science: Today and Tomorrow', *Mind and Society* 1, no. 1 (2000): 25–39.
[530] Grüne-Yanoff, 'Bounded Rationality', 545.

> Complete 'realism' is clearly unattainable, and the question whether a theory is realistic 'enough' can be settled only be seeing whether it yields predictions that are good enough for the purpose in hand or that are better than predictions from alternative theories.[531]

But if this is the case

> The question arises why these theories insist on rationality, now in bounded form, at all. Why not just construct a causal theory of behaviour, which allows any kind of irrationality, if needed?[532]

The answer to this is simple: theories of bounded rationality emerge only because there *is* a difference between precise Bayesian or EUT math, and the everyday human decision maker and their risk and probabilistic decisions. Rationality is useful to expel the bounded-ness of human consideration and action, but it is not the only way to reveal our limits. Thus, if we are considering the applicability of our ethics, EUT and Bayesian analysis are only helpful in that they reveal possible faults when considering our decision making in terms of the probabilistic or risk assessment of a particular scenario occurring.

It follows that I disagree with the common statement in DT literature that standard or classical DT (e.g. EUT and Bayesian analysis), is normative. The Bayesian structuring or EUT probability is a descriptive consideration of normative beliefs. On this level normativity is a mute token, subordinated to mathematical relation(s). The beliefs, or the normativity, are just the values along for the descriptive ride, and to insist otherwise is to breach the is-ought distinction.

10.4 Multi-Criteria Decision Analysis

Multi-criteria decision analysis (MCDA) is a broad term, covering many different methods for assessing decisions with multiple criteria, for ranking, selecting and/or comparing

[531] Friedman, *Essays*, 41.
[532] Grüne-Yanoff, 'Bounded Rationality', 553.

different alternatives. The three common approaches to be found in the literature are (i) *utility function theory*, (ii) *outranking relation theory*, and (iii) *decision rule theory*:[533]

(i) Utility-based theory includes methods synthesising the information in a unique parameter.

(ii) Outranking relation theory involves methods based on comparisons between pairs of options to verify whether 'alternative *a*, is at least as good as alternative *b*'.

(iii) Decision rule theory originates from the artificial intelligence domain and it allows deriving a preference model through the use of classification or comparison of decision examples.

Utilising DT when considering environmental issues is not new,[534] and MCDA is most typically used. This is because one significant advantage of MCDA is that it does not apply reductionism to values so as to compare them along a single scale.[535] A further advantage of adopting MCDA is that it enhances the transparency of consideration and comparison of values, and it allows for potential repeatability in decision making.[536] If aggregation and weighting are done well, then the decision making algorithm could be reapplied in like-for-like scenarios. However, because these steps of aggregation and weighting require so much detail as to not inadvertently favour one option over another, MDCA has a significant disadvantage:

> In the majority of the available assessments, the selection of MCDA method is dependent on the familiarity and affinity with the approach rather than on the decision making situation under consideration.[537]

Sarkar's MCDA method, as developed for environmental decision-making, includes input from the environmental sciences at both the initial comparative stage and the hierarchal stage, as outlined in Table 10.1.[538] Sarkar's approach is to incorporate ethics/values into

[533] Marco Cinelli , Stuart R. Coles, and Kerry Kirwan , 'Analysis of the Potentials of Multi Criteria Decision Analysis Methods to Conduct Sustainability Assessment', *Ecological Indicators* 46 (2014): 140.
[534] Colyvan, 'Environmental Philosophy', 96.
[535] Peterson, *Introduction*, 172.
[536] Sarkar, 'Multiple Criteria and Trade-Offs', 535.
[537] Cinelli, Coles and Kirwan, 'Analysis of the Potentials', 146.
[538] Sarkar, *Environmental Philosophy*, PAGE?

MCDA, and then to use EUT to reveal probabilistic chances of a particular policy options success; though the ultimate decision on policy should be left to 'social policy'.[539] (It is notable that with this proviso he leaves open the option of ignoring their input altogether.)

In other environmental applications of MCDA, sustainability is categorised into weak or strong perspectives to deal with this issue in the early phases of the analysis:

> From a weak sustainability perspective, different forms of capital (e.g. financial, human and ecological capital) are substitutable. For example, the loss of a rainforest ecosystem (ecological capital) may be offset by the financial capital gained from the development erected in its place. From a strong sustainability perspective, this is not the case.[540]

This is not my suggestion. As I see it, MCDA is not a matter of picking the 'best'' ethic**,** but of displaying their respective possibilities for real operational success. The idea is to utilise the diverse qualities of MCDA, but not subordinate the final decision to its findings. EUT or Bayesian analysis is helpful in so far as it deploys statistical rationality in decision analysis. However, the use of these techniques should not directly state or justify a particular ethic over another, for this is a breach of the is-ought distinction.

While Sarkar argues that any descriptive account of DT (e.g. prospect theory, and those theories concerning how we actually behave) is irrelevant, because he believes that EUT is a normative gambit,[541] I have argued that EUT and like analyses are mathematically descriptive uses of 'normative tokens'. Descriptive accounts are more relevant than Sarkar argues to ensuring the decision reached reflects a weighting of *oughts*, and can be added alongside the inputs listed in Sarkar's table as 'inputs from the biological and cultural sciences' (see Section 10.5). It is here that the opportunity arises of adding these human sciences to the MCDA framework.

For example, our descriptive account in terms of bounded rationality, as revealed by such EUT or MCDA techniques, may in fact present such probabilities via an MCDA. It

[539] Sarkar, 'Multiple Criteria and Trade-Offs', 535.
[540] Rowley et al., 'Aggregating Sustainability Indicators', 30.
[541] Sarkar, *Environmental Philosophy*, 81.

presents to an audience of human brains *in a certain way*, and then presents them with a stimulus as a cue to favour a specific choice, (e.g. people are likely to favour statistical probability – high chance of success per embedded ethic X – over low probability of embedded ethic Y —per heuristic A or neural causation B).

The descriptive account seems vital if we are not to let EUT/Bayesian results skew the inherent ethics (the normative tokens). Furthermore, rational theories in the mathematical sense may be improved or adapted in that field by such sciences. It would take a mathematical thesis to expound theories like Greene's into a revised EUT/Bayesian form. While we wait for this, the descriptive account can be incorporated into the MCDA analysis that treats and assesses each ethic, with the inclusion of causal descriptive facts concerning our decision making, in much the same way as we consider the science of environmental degradation to be relevant to options for environmental preservation. With this addition of inputs (marked with a "?" in Table 10.1), we achieve further realism about the would-be applicability of our environmental ethics with respect to Paradigms A to D in Table 10.1.

Table 10.1 Inputs from the Environmental Sciences

Paradigm	Environmental Science
A. Assessment of the status of an environmental problem	The relevant science would provide a model that predicts what would happen if no action is taken. In principle this appears straightforward. However, prediction in the ecological and social sciences depends on so many variables that precise prediction of outcomes is often impossible. Hence the importance of quantitative uncertainty assessments.
B. Design of policy options	The environmental sciences help in designing scientifically viable policy options, which are the alternatives available to decision makers. Once again, accurate prediction is an issue that must be explicitly treated.
C. Assessment of the consequences of each policy option	The environmental sciences should provide tools to predict the outcome of each policy option. Note that taking no action is also a policy option: in that sense the status assessment stage above can be subsumed under this stage. If all consequences of policy options are judged unsatisfactory (which is not a purely scientific issue), then we must return to the design stage and formulate more policy options.
D. Estimates of the uncertainties of each assessment	A highly desirable policy outcome may have a very low probability. A somewhat less desirable outcome may have a much higher probability. We may (but need not) choose the latter alternative. It is not the task of the environmental sciences to decide such differences; it is rather the choice of social policy. However, what the environmental sciences should do is specify how probable each policy outcome is, preferably quantitatively.

Source: Sahotra Sakar.

10.5 Towards 'Design Features' for Policy Creation and Assessment

10.5.1 Introduction

We can extend Table 10.1 by the careful inclusion of inputs from those causal-reductive sciences explored in this thesis. These inputs bear on the assessment of the status of an environmental problem, the design of policy options, the assessment of the consequences of each policy option, and estimates of the uncertainties of each assessment.

Obviously I cannot provide a full account of all possible applications of such sciences to all possible EEEs. Instead, I propose a range of questions that arise on the back of the sciences as they may bear on these matters. This is the topic of the next Section.

10.5.2 Format of Questions

Many of the causal factors revealed in this research apply not just to the populace to which the policy is intended to apply, but also to the individuals or groups making such policy. This point is important where an agreed set of values is required.[542] I begin by proposing some definitions:

P(CE) = A Current/Existing Policy with specified Actions, and Embedded Environmental Ethics (EEE).
P(AE) = A New Policy's Proposed Actions, with EEE.
X,Y = Areas of reductive sciences, assumed as relevant by the question.
Z = Environment Current
Z1 = Environment Intended by EEE

Applicability Potential = P(AE) actions applied in environment Z, where P(AE) is casually affected by (X,Y), such that Z1 is more or less likely realised.
Applicability Fail = P(CE) actions applied in environment Z, are casually affected by (X,Y), such that Z1 is not being realised.

[542] Rowley et al., 'Aggregating Sustainability Indicators', 25.

Table 10.2: Format of Questions in the Context of Paradigms

A. Assessment of the status of an environmental problem	**Applicability Fail** = P(CE) actions applied in environment Z, are casually affected by (X,Y), such that Z1 is not being realised (as evidenced by Environmental Science).
B. Design of policy options	**Applicability Potential** = P(AE) actions applied in environment Z, where P(AE) is casually affected by (X,Y), such that Z1 is more or less likely realised.
C. Assessment of the consequences of each policy option	**Applicability Potential** = P(AE) actions applied in environment Z, where P(AE) is casually affected by (X,Y), such that Z1 is more or less likely realised. **Applicability Fail** = P(CE) actions applied in environment Z, are casually affected by (X,Y), such that Z1 will not likely be realised.
D. Estimates of the uncertainties of each assessment	**Applicable Probability**: Assign, where possible, respective probabilities determined from A (precedent), B and C (predictive). Utilise applicability probabilities in MCDA, along with environmental science inputs based on proposed policy, to weight policy options.

The challenges entailed by the use of this format are as follows:

- P(CE) infers a probability distribution. Such policies are meant to be defined over either a specific action ("Stop burning coal") or a conditional plans ("If warming goes over 0.5C, stop burning coal")
- A highly desirable policy outcome may have a very low probability. A somewhat less desirable outcome may have a much higher probability. We may (but need not) choose the latter alternative.
- It is not the task of the environmental sciences to decide such differences; it is rather the choice of social policy. However, what the environmental sciences should do is specify how probable each policy outcome is, preferably quantitatively.
- How is the low probability of a highly desirable policy is to be specified? Is the probability a matter of 'political' probability—in which case how could ecological

or environmental science have anything useful to say on this 'probability'? After all, such a claim is a political one, not a scientific one.

Environmental science itself will not decide the chosen policy, because it arises from social policy, as Lynch and Sarkar make clear. However, the reductive human sciences reach further into the realm of those interacting brains that formulate social policy, and further into the cause and effect of decision making and behaviour, than the environmental sciences themselves. For instance, such sciences may let us see why 'a highly desirable policy' has, given the nature of our brains, a low probability of being pragmatically successful. This view of our causal behavioural response is important and, when it comes to an ethic of sustainability, should sit alongside the environmental sciences.

How is the low probability of a highly desirable policy to be specified? I answer by separating the highly desired end result of the EEE embodied in a policy, from its application.

For example, consider a policy option with EEE, which has a desired end result specified by the EEE. However our understanding how human brains work (X,Y) reveals that this result is unlikely (even impossible) to operationalise on the required scale**.** At this point we can ask what alternative policy construction options might be better suited to delivering the desired outcomes. Given the relevance of (X, Y) we might be able to see such options more clearly. This way we *extend* traditional environmental ethics from moralism to policy without breaching the is-ought distinction, or putting our ethics second in decision analysis.

10.5.2.1 Moral Evolution (Inclusive of Sections 3-4)

Beginning with Section 4.2 concerning the general principles of evolution and evolved moral brains, we might *initially* ask:

Paradigm B or C

- Does the policy take into account the nature of the cooperative suite of neural functions humans evolved via group selection? And does it do so in a way that may influence its potential effectiveness? (E.g. taking into consideration, or directly

leveraging, moral safeguards like retribution, punishment, guilt, generosity and competitiveness.)

- Is the policy option subject to 'subversion from within', given the nature of individual selection?
- Does the policy option make sufficiently clear the costs and benefits of cooperation in order to engage the moral brain**?**
- Does the language of the policy option take into account the variance in cultures to which the policy may be applied? (While the evolved moral brain is consistent, the cues that elicit which parts of it may be influenced by such cultural differences.) Furthermore, does the language of the policy option align or entail the physical circumstances it gives rise to?
- Are decision makers influenced by these factors in a way that might hinders a more fitting and effective policy from being constructed? (E.g. Can we assess the current political environment openly?)
- Does the policy option actively engage evolved social behaviour given Churchland's four-dimensional scheme? (a. caring; b. recognition of other's psychological states; c. social problem solving concerning resource distribution; d. learning of social practices via negative reinforcement, trial and error, imitation, analogy etc.)
- Does the policy option actively engage Mikhail's evolved Universal Moral Grammar? Along similar lines, does it act against Sripada's view of content nativism?
- Is the policy option realistic in terms of the environment in which it is to be applied, given that the brains of children and young adults have a large degree of flexibility in terms of their moral compass, which is cemented by the nature of these environments?

Paradigm D

- Can we reduce or better predict uncertainty in specific cases in which it is clear that such behavioural outcomes are likely to occur?
- Does a specific cultural mechanism influence or alter this uncertainty about the prior question?

Paradigm A

- Has the language of the policy option hidden or misled us about the actual operations of our moral brains, and has this itself fed into our destructive behaviours?
- Does the policy language automatically cue counter-effective moral inclinations?
- Is the behavioural cause of the environment problem going to change or alter (e.g. increase in activity) given what we know about our evolved cooperative capacities as 'moral creatures'?

10.5.2.2 Evolutionary Psychology (Section 5)

Paradigm B or C

- Are there any known specific modules (i.e. cheater detection) that may influence the outcome or effectiveness of the policy designs (Paradigm B) or their potential consequences (Paradigm C)?
- Does the policy have provision to punish cheaters, and to sufficiently punish those who fail to punish cheaters? If the cheater detection module is not satisfied/is continuously-elicited, cooperation will fail (e.g. obviously abusing a welfare system over and again creates division between the cheaters, those not punishing them sufficiently (the government via its policy), and cheater detectors who fund the welfare system (taxpayers)).
- Does the policy option pertain to deontic logics ('If you give me Q, I will give you P', and/or social contracts, 'We both benefit if we trust each other' etc.), which may, as design features, either aid or disrupt the intent of the policy?
- Are there any specifically proven heuristics that may influence the outcome or effectiveness of the policy design, or their potential consequence? (In answering this question we must think about the environment in which the policy will be applied.)

(a) Recognition heuristic: when making a judgment about two items, the more easily recognised item will be considered to have a higher value (i.e. will the policy option be ignored in lieu of a more recognised option).

(b) Fluency heuristic: if both alternatives are recognised but one is recognised faster, then infer that this alternative has the higher value with respect to the criterion.

(c) Default heuristic: If there is a default (in a social construct), do nothing about it (it is argued such defaults are set by institutions) (i.e. Is the proposed policy option leaving open the gate as 'to do nothing about it'? Is there established institutional opposition that also needs to be considered for the policy to be successful?)

(d) Representative errors: Does the policy being considered inherently arise from such errors on the part of the policy makers?

Paradigm A or D

- Are there any current modules or heuristic rules that are causing a policy option to fail by generating adverse behavioural responses, because the policy option had not considered them? Is there any information present in the current environment, which may have changed since the policy was created, which has been unaccounted for in the existing policy?
- Can we reduce the uncertainty of a proposed policy option by the Applicability Fail precedent established in an existing policy?

10.5.2.3 Cultural Evolution (Section 6)

Turning to Section 6.3, we might perform a Culture Impact Assessment in order to understand those cultural influences that are relevant to a policy option. This approach specifically incorporates the prior sections:

Paradigm B or C

- If we test a culture's replication dynamic (likelihood/frequency of its adoption), can we then reduce uncertainty of particular policy option on the basis of its cultural acceptability?

  $\Delta p = p(1-p)[X-Z]$, where: ΔP = frequency of individuals who adopt the culture, in that it modifies their behaviour; X is strength of content bias (*the psychological module processing or affected by the introduced culture*); Z is the cost to an individual who acts on the cultural information in terms of passing it on to another (its cultural fitness).
- Can we better design policy options and assess their consequences by forecasting their possible replication dynamic based on precedent cases? Can we identify possible transmission modes for the proposed policy option? (e.g. cultural variation, cultural selection, cultural drift.)
- Is the policy option addressing an already present meme/culture/idea? Can we use a Cavalli-Sforza model to determine whether their differential probability of adoption by others will increase or decrease?

Paradigm A or D

- Is an existing policy option failing because it has a low frequency of adoption? Is this lack of adoption primarily due to one of the reasonings (content biases) found in Sections 10.5.2.1 or 10.5.2.2?
- Can we perform a fuller Darwinian analysis on specific cultural factors relevant to the proposed policy, following the steps already cited as below:

  a. Draw up a model of the life history of individuals;

  b. Fit an individual level model of the cultural transmission process to the life history;

  c. Decide which cultural (and if relevant genetic, module or heuristic capacity) variants to consider;

  d. Fit an individual level model of the ecological effects to the life history and to the variants;

  e. Scale up by embedding the individual-level processes in a population; and

  f. Extend over time by iterating the one-generation model.

10.5.2.4 Cognitive and Behavioural Neuroscience (Sections 7–9)

Paradigm B or C

- Does the policy option initiate motivational value proxies with regard to dopamine behavioural responses?
- Does the policy option consider previously established reward connections within the targeted population, resulting in behaviours contrary to policy design and success? To this end, does the policy option provide a substitute value proxy to ensure no unexpected and adverse substitute occurs?
- Given the apparent transitivity in decision-making valuation at the neural level, what is the likelihood of policy success/adoption, given the decision maker will equate policy factors and existing factors/alternatives under a common scale?
- Using a sample group, is it possible to test the value of a policy option in terms of choices made by constituents expected to implement it, ideally via fMRI to see if the choices made in favour of it obtains a threshold firing rate in light of alternatives (the value of other available alternatives). (This may sound expensive, but testing 20 people via fMRI for a policy applying to >2million people, is arguably quite efficient, especially if the policy option could have adverse outcomes if its optionality is misguided.)
- Is the policy option likely to be effective or ineffective given the nature of the emotional structure described in the SMH, SRR, SEC or EFEC frameworks? Is a policy option relying on a processing or representational account of emotional control?
- Given DPT finds that removing time pressure and encouraging deliberation increases utilitarian judgement, does the policy option consider this outcome?
- Given DPT finds that argument strength and spatial awareness have effects on whether ‘manual’ or ‘automatic’ brain processes are selected, does the policy option consider these factors?
- Does the policy appreciate the distinction of causal priority found in the emotional brain illustrated by DPT?

- Given Louie, Khaw, and Glimcher's[543] test findings on human subjects (the 'neural representation of the value of an option is explicitly dependent on the value of other available alternatives'[544]) can we measure the value of alternatives via fMRI to ensure it represents the actual normative ethical value?

Paradigm A or D

- Is the policy option being made under stress, and therefore discounting future rewards in favour of smaller immediate rewards?
- Is the policy option being made without sufficient deliberation, leading to less utilitarian decision making?
- Given people who are more empathetic, or induced to be more empathetic, give more deontological responses, is this factor skewing the policy construction in a capacity that effects the application of the ethic at incorporated?
- If, for example, a presenter induces mirth among the policy decision makers (recall mirth is the positive emotion associated with humour, which has been shown to increase consequentialist responses), has this capacity skewed the decision being made?
- Is there any distraction present in the environment of the decision maker that has reduced a consequentialist decision from being made?
- Is the environmental destruction in a specific region a result of an environmental factor elicited by the policy option, whereby that environmental factor, unaccounted for by the policy option, cues the automatic brain per Greene's DPT?
- Given the stark evidence of VMPFC damage and its effect on decision making, should the policy option creators be tested to ensure no such damage is present?
- Can Greene's DPT help us predict the probability of success of a policy option based on the above facts in the policies' specific case?

[543] Louie, Khaw, and Glimcher, 'Normalization'.
[544] Ibid., 6139.

Chapter 11: Summary and Conclusions

11.1 Thesis Overview

I have identified context-specific methodological issues on a range of traditionally controversial areas for ethicists – including utopianism, reductionism, determinism, naturalism, evolution, and pragmatism – and on the use of incomplete and multileveled sciences. I have developed a methodology for utilising the sociobiological, reductive, natural sciences so that we might exploit the ways in which they may (and may not) assist in the effective application of environmental ethics.

On the basis of this methodological position I analysed the relevant reductive sociobiological sciences – including evolution and genetics, moral evolution, evolutionary psychology, cultural evolution, and cognitive and behavioural neuroscience – with an eye to effective policy implementation.

Having developed a methodology and having analysed the sciences it utilises, I then made the connection to the decision-making and policy construction frameworks, DT and MCDA. This culminated in the presentation of multiple questions and considerations for further research, all of which are embedded in my underlying concern to effectively manage the sustainability challenge of the Anthropocene epoch.

11.2 On the Methodological Tenets

First, I noted that utopianism presents a very real threat that can be all too easily disguised in naturalistic visions of human organisation and self-betterment. I argued that, even in the broad sense captured by Schofield (Section 2.1),[545] the suggested use of facts of a causal naturalist kind was not utopian. This was because science was not itself given an ethical status, nor was it used to place one particular set of ethical instructions over another. If one

[545] Schofield, *Plato*, 234.

had a well-reasoned ethical theory in a domain, this kind of causal science about decision-making was clearly relevant when assessing the social applicability or operationalisability of this ethic.[546]

Second, I argued that reductionism may be used to expound causal theories of practical relevance to the application of ethics. This did not mean supporting eliminative reductionism, nor succumbing to the naivety of a purely materialist or scientism outlook. We should take the mereological fallacy seriously, but only to the point where a reductive ensemble can meaningfully produce a causal theory relevant to the application of our ethics. This was demonstrated with a mechanistic sketch of the relationship of manipulability.[547] Although there remain many issues with the realisation of mechanistic causality, these should be used to dogmatically dismiss the careful use of this reductionism.

Next, I argued that UCD, stated simply as 'X future is a product of the prior state of the universe Y', was a plausible implication/result of causal reductive naturalism, but that that shouldn't mean obliteration of the reason-cause distinction. A basic reason was that scientific explanations give us the *causal reasons* for the occurrence or non-occurrence of phenomena. We cannot have our cake (seriously discuss the findings and implications – including practical implications of our new sciences) and eat it (by rejecting UCD).

Fourth, a commitment to methodological naturalism did not mean reducing values to facts (though that this or that is a value, or valued by us, was, of course, just another fact).[548] As intimated in the previous paragraph, and as a constitutive assumption of this thesis, I assumed a distinction between theoretical reason(ing) concerning 'what is the case?', and practical reason(ing) concerned with 'what shall we do?'

Having argued for a naturalism that respected this distinction, I insisted this did not mean the two modalities were insulated from each other. I argued that facts and values can come together fruitfully if we follow Bernard Williams' suggestion and considered issues of what we (human beings constituted as we are through evolutionary history) 'can and cannot' do. Obviously, as Aristotle long ago pointed out criticising Plato for thinking

[546] Traer, *Doing Environmental Ethics*, 163.
[547] Craver and Robins, 'Biological Clocks', 52.
[548] Norton, 'Facts, Values, and Analogies', 65.

people might live flourishing lives without special kin or family connections, such facts bear directly on the applicability and operationalisability of ethical policies.

Given all this, I concluded that we must accept a naturalistic biological basis for conscious activity if we are to leverage its causality. This did not demand we identify the experience itself with this basis, only that it was the causal origin of that experience.[549]

I then constructed a simple epistemology for truth claims in terms of observational evidence (Section 2.4.3) that allowed us to rule out any talk of supernaturalism or mind-body dualism.

In general I argued for an environmental pragmatism, whereby one 'develops strategies by which environmental ethics can contribute to the resolution of practical environmental problems'.[550] Though I did this in the context of insisting that this project can only work if we did not lose sight of those intrinsic feelings we have about nature. These feelings are facts. Thus we need not, and should not be instrumental about our intrinsic values.

Finally, I considered the problem that the sciences explored came from various emerging research fields and were as yet incomplete. No single account explained moral or ethical behaviour. What was required was a sharp lens across the spectrum that pulled together all those aspects relevant to a specific inquiry. This meant I rejected 'ruthless reductionism' for a 'mechanistic reductionism', looking to system-level explanations of brain composition.[551]

11.3 On the Sociobiological Fields

Beginning with evolution and genetics I discussed the various pitfalls to be avoided in describing and applying evolutionary theory. The most controversial positions I took involved accepting the notion of 'levels of selection', and agreeing with the potential existence of 'group selection'. Not quite so controversial was my claim that while some

[549] Searle, 'Consciousness', 1.
[550] Palmer, 'An Overview'.
[551] Bechtel, 'Molecules, Systems, and Behaviour', 14.

early claims were much inflated, genes remain an important part of the causal story of our decision making, and central to any adequate multi-level causal story.

Having discussed the basics of evolution and possible misapplications, I addressed the evolution of morality, starting with what constituted morality and moral judgement. While neuropsychologists may have been overly ambitious, or in cases simply careless, when it came to these questions, I argued that what they did say was clearly enough to see the relevance of their inquiries and findings. The common approach in the literature was to stress the idea of *prohibition*, so as to distinguish moral behaviour from merely altruistic behaviour. This behaviour (at the very least, central to anything we call morality) was then explored in terms of the evolution of cooperation and of attachment and bonding. While much of the argument was arguably of the 'just so' form, there were no serious counter proposals to this reverse engineering strategy, only smaller debates concerning the order in evolutionary history.

Before investigating these causal proximities further, I considered the literature on nativism (the origins of which were discussed in reductionism and causation in Chapter 2), and the difference between emotional moral decisions and that born of reason. The central question was: how much of our moral and ethical capacity is innate, and how much is a product of reasoning? Hauser, Mikhail, Pinker and DeWaal argued for an evolved innate moral suite, Prinz and Sripada argued the other way.

This debate deepened when framed in terms of emotion *v.* reason as the origin of our moral construct. Despite the significant points raised by these arguments, they were not sufficient to dismiss the idea of neural functions or modules directing our moral thought and action, whether in the register of emotion or reason. What they did was lead us to that evolutionary psychology that attempts to understand these neural capacities.

Distinct from cognitive psychology, evolutionary psychology explains behavioural capacity and origin from particular, often supposed, environmental accounts of our developmental history via evolutionary mechanisms as they operated over geological time.[552] This theory had its roots in Jerry Fodor's 'massive modularity thesis'.[553]

[552] Viren, *Evolutionary Psychology*, 32.
[553] Scott M. James, *Introduction*, 18.

Cosmides and Tooby began such an investigation in 1985, building their 'social contract theory'[554] concerned with the type of reasoning that developed through situations of social exchange. At the heart of their theory was the notion that the brain is analogous to a computer.[555] The first challenge to this theory targeted the compartmentalisation of the brain itself.[556] The general response was that one should not think of a module as a localised area of the brain; there were different types of modules that consisted of multiple brain areas. The most researched of these was the Cheater Detection module. This was Cosmides and Tooby's initial module,[557] born from such experiments as the Wason Task,[558] and later expanded to account for other aspects of our moral behaviours.[559] Further support for this approach arose in the form of mathematical models, which show that cooperation may evolve and remain stable if individuals punish both cheaters and those who fail to punish cheaters.[560]

The stimuli that triggered such modules in social exchange did not need to occur eternally; they might occur in the imagination,[561] or meta-representations.[562] This meant that deliberative reasoning was not insulated from the influence of domain-specific evolved programs.[563] Evolutionary psychologists found that reasoning took the form of two systems (though there are likely more): one for social exchanges, and one for precautionary rules. The former contained a myriad of social contract algorithms that might be triggered by the specific use of certain words, or just actions.[564] Neuroscience correlated with this theory, showing different brain areas were utilised for social exchange reasoning and that for precautionary rules.

Peter Ulric Tse had shown how 'cross-module binding' might explain our capacity for analogical reasoning (central to much moral reasoning).[565]At the heart of Tse's argument

[554] Cosmides, 1985.
[555] Tooby and Cosmides, 'Conceptual Foundations'.
[556] Viren, *Evolutionary Psychology*, 34.
[557] Cosmides and Tooby, 'Cognitive Adaptations'.
[558] Wason, 'Reasoning'.
[559] Cosmides and Tooby, 'Cognitive Adaptations'.
[560] M. D. Hauser, *Moral Minds*, 432, quoting Boyd and Richerson, *Not by Genes Alone*.
[561] Boyer, *Religion Explained.*
[562] Leslie, 'Pretense and Representation'.
[563] Cosmides and Tooby, 'General Deontic Logic', 56.
[564] Ibid., 113.
[565] Tse, 'Symbolic Thought', 281.

was the idea that through binding humans developed the ability to form associations across previously encapsulated classes of information.[566] Tse's point aligned with the intent of this thesis:

> If there is a kind of tokenisation that does operate with racist sexist, and discriminatory behaviour, then the theory may be helpful in developing methods in moral education that aim to disrupt the processes which lead to such attitudes.[567]

I looked next looked at the field of heuristics. If we did have a cheater detection module, this module, according to Gigerenzer, presented the downstream 'intuitive search rule'; a moral heuristic that sought to reveal if one had been cheated in a social contract.[568]

Evolution has not just been applied to the formation of our brains but to the very ideas they produce. Richard Dawkins' cultural evolutionism hypothesised a 'cultural replicator' that acted on the level of ideas, as did genes on the biological plane. He called such replicators 'memes'. However, for a number of reasons (beginning with controversies at the purely biological level) this strict replicator analogy wasn't essential to cultural evolution. CE can happen much faster than biological evolution, and culture can constitute a change in ecological conditions that, in turn, may alter selection pressures. Successfully modelling cultural transmission is, then, an ongoing challenge. Here Boyd and Richerson's 'Darwinian analysis', and Henrich's 'replicator dynamic equation' offered methods for tracking cultural evolution and its effect on individuals. Such models, adequately promise a pragmatic toolkit for considering the effectiveness of an EEE.

With an understanding of evolution, the evolution of morality, evolutionary psychology, heuristics, and cultural evolution, the next stage of the reductionist enterprise was to engage functional composition and causality through the lens of cognitive and behavioural neuroscience.[569] The requirement was that any proposed causal theory should have a reasonably coherent evolutionary construct.[570]

---

[566] Ibid., 276.
[567] Wallace, 'Morality', 310.
[568] Gaissmaier and Gigerenzer, 'Heuristic Decision Making', 478.
[569] Bechtel, 'Molecules, Systems, and Behaviour', 34.
[570] Wood and Grafman, 'Human Prefrontal Cortex', 139.

I looked first at issues with scanning for moral circuitry, especially Berker's scepticism concerning what fMRI scanning can be said to reveal.[571] Taking Berker's points on board, I argued in favour of Joshua Greene's position as one that aligned with the requirements of developing a mechanistic sketch and eventual relationship of manipulability. With fMRI limitations in mind, I concluded, first, that we must be wary of neuroscientists expressing false casual reductionism through absolute language based on their clearly limited experiments. My second point was that we should appreciate the extent to which these experiments fell short of capturing the full experience of moral decision making. Still, neither point meant denying that such experiments may reveal useful causal information concerning aspects of our decision making, even if the causality was blunt, simplistic, and narrow.

Next I presented an overview of neural connectivity and function, before moving onto neural decision making and the motivation, valuation and choice models on offer. With the former, I argued for the importance of Landreth's theorising on motivational control. My account of valuation drew heavily on Padoa-Schioppa and Assad's work on capuchin monkeys' decision making.[572] They found the monkeys made an evaluation of the food pairings placed in front of them based on the quantity and quality. That suggested these subjective values were actually encoded in the striatal neurons where, presumably, they were engaged by the choice module. It was shown the encoding of subjective values in the case of valuation were invariant, determined by weighting on a single scale, with choice a matter of neuron firings in a set of areas in the parietal lobe, with the subject value chosen the one that had a firing rate that exceeded a local threshold. This process was theorised as *decisive normalisation*.[573] Before this it was thought that choice was context-independent.[574] While this research did not examine multi-attribute decision-making involving the PFC, it seems clear this process will yield significant insight into human decision.

I then considered a number of psychological studies, only a few with fMRI support, concerned with emotional decision making; in particular those which measured the effects

[571] Berker, 'Normative Insignificance of Neuroscience', 293.
[572] Padoa-Schioppa and. Assad, 'Neurons in the Orbitofrontal Cortex'.
[573] Louie, Khaw, and Glimcher, 'Normalization', 6139.
[574] Ibid., 6142.

of stress and mood across a broad range of scenarios.[575] Procelli and Delgado showed that stress intensified risk seeking when choosing between possible loses, but resulted in a risk adverse approach when choosing between potential gains.[576] Youssef et al. showed that stress decreased the likelihood of making utilitarian judgements in personal moral decisions.[577] Nguyen and Noussair, using face scanning software under a number of test conditions, found that positive emotions (not being stressed) correlated with greater risk taking.[578] Kimura et al. showed that stress exaggerated the tendency to discount future rewards in favour of smaller immediate rewards.[579] Harle et al. showed people were more likely to reject unfair offers when they were in a sad mood.[580] Lighthall et al. (here with fMRI evidence) showed that under stress males were more risk adverse than females.[581]

Having assessed valuation, choice and emotional decision making in the areas of the brain not involving the PFC, I turned to studies involving the PFC. Here the DLPFC, VMPFC, and the Amygdala, were the most significant areas.[582] Having looked at medical studies of persons with brain damage, it was clear that if the PFC were damaged we could expect certain behavioural effects. Damage to this area did not affect language, memory or IQ, but rather 'ambition, empathy, foresight, a complex personality, a sense of morality'.[583] Koenigs et al. found that patients with VMPFC lesions were significantly more likely to endorse utilitarian responses to moral dilemmas.[584] Similarly, Eslinger and Damasio observed a distinct lack of moral behaviour from patients with lesions, but surprisingly a normal moral response when presented with moral reasoning tasks;[585] a finding that supported the idea that the VMPFC plays a significant role in selecting emotions during moral reasoning tasks.

These studies require we take seriously the idea that utilitarian and deontological moral decision making are constituted by neural pathways within the PFC. Whilst such studies are preliminary, the causal information being revealed is worth serious attention.

---

[575] Phelps, Lempert, and Sokol-Hessner, 'Emotion and Decision Making', 273.
[576] Porcelli and Delgado, 'Acute Stress'.
[577] Youssef et al. 'Stress Alters'.
[578] Nguyen and Noussair, 'Risk Aversion'.
[579] Kimura et al., 'Biological Effects'.
[580] Harle et al. 'The Neural Mechanisms'.
[581] Lighthall et al., 'Gender Differences'.
[582] Greene, *Moral Tribes*, 123.
[583] Ramachandran, *Tell-Tale Brain*, 21.
[584] Koenigs et al., 'Damage'.
[585] Eslinger and Damasio, 'Severe Disturbance'.

To see how we might make use of this information, it was necessary to consider the nature of the PFC construct itself, and I considered a number of representational and processing theories. Whilst most of these theories fell short of adequately meeting Wood and Grafman's criteria,[586] a hybrid theory seemed most likely for reasons I discussed.

At this point I was in a position to assess the more relevant PFC moral theories. I began with Damasio's SMH, which argued that decision-making could not be a product of cognition alone, and that automated/involuntary emotional networks determined much of our moral decisions. The core idea behind Damasio's theory was that these emotions were 'labelled' in the VMPFC; thus Damasio theorised that damage to the VMPFC, which arguably linked internal representations with somatic (nervous system) responses, explained specific changes in moral decision making and behaviour. I then looked at James Blair's social response reversal (SRR) theory, which suggested a hierarchical relation between cognition and emotion, with specific models suppressing aggressive behaviour. If these models were damaged, then the various types of aggressive behaviour resulted. Koenigs et al. and Tabber-Thomas have since found evidence supporting this theory.

Wood and Grafman's SEC framework went one step further, arguing that event knowledge was actually stored within the PFC, and that damage to this area of the brain was damage to built-up/reinforced moral knowledge. Unfortunately the theory had little to say on how the PFC stored this information.

The EFEC framework took elements from all of these theories in the form of a binding of information within the PFC through an integrated processing mechanism.

Left with multiple possibilities for a theory of moral cognition, I turned to Greene and DPT and MPT.

DPT rested on the idea that both cognitive/reasoning 'manual mode' and emotional processes 'automatic mode' determine moral judgment, and that which is selected depended on the environmental cues present.[587] After performing modified Trolley-based experimentation and fMRI testing, Greene argued that the emotional systems of the brain

[586] Wood and Grafman, 'Human Prefrontal Cortex', 139.
[587] Greene, *Moral Tribes*, 202.

were activated by the potentiality for personal engagement, and the actual proximity of that engagement. Paxton, Ungar, and Greene then demonstrated that reflection increased the likelihood of utilitarian moral reasoning determining a decision. The key determinates of reasoned reflection were sensitivity to argument strength, and extended temporal duration.[588] Latter experimentation by Greene let him to MMT, which accounted for how emotional systems can overrule the cognitive, manual mode.

Taken together DPT and MMT yielded a more substantial mechanistic sketch than the other theories discussed, and one that could be useful to the application of environmental ethics.

11.4 On Decision Theory and Policy Construction

The aim of the previous chapter was to apply the findings of the previous chapters to environmental decision making and policy construction. CBA and the PP were the obvious frameworks to initially consider. Their respective failures and successes were already well established in the literature, and I sought only to distinguish my position here. In this effort I was careful not to breach the is-ought distinction, leveraging the methodology in chapter 2. Although descriptive accounts have been provided by Elinor Ostrom and, in a different capacity, Kahneman and Tversky, the approach I developed is missing from the field of decision analysis. The challenge then was how to incorporate sociobiology into traditional frameworks, especially given the early developmental state of the sociobiological sciences. With so many questions still needing to be answered, the initial task became designing a helpful categorical format for their input relevance in a MCDA.

To begin doing this I looked more closely at bounded rationality, and, in particular, Bayesian decision frameworks beyond classical DT and their internalist and externalist perspectives. While these do not, and should not be interpreted as setting the ends of our deliberations, they are crucial in so far as they reveal possible faults our probabilistic reasoning and risk assessment. I considered the question as to whether such accounts could be aligned with sociobiological causation in a grand MCDA: the possibility for an ethical

[588] Paxton, Ungar, and Greene, 'Reflection and Reasoning', 164.

decision analysis framework that leveraged mathematical rationality *and* sociobiological theory.

Such an effort would undoubtedly have to wait for an appropriately qualified interdisciplinary team, but what I could offer was considered questions stemming from the sociobiology examined, presented in a format that would align with the paradigms set out by Sahotra in Table 10.1. My argument is that the causal information found in these emerging areas of sociobiology ought to be available to improve our environmental policy construction and decision making.

Bibliography

Abend, G. 'What the Science of Morality Doesn't Say About Morality'. *Philosophy of the Social Sciences* 43, no. 2 (2013): 157–200.

Aizawa, K., and C. Gillett. 'Levels, Individual Variation, and Massive Multiple Realisation in Neurobiology'. In Bickle, *Oxford Handbook*, 540–581.

Aldred, Jonathan. 'Justifying Precautionary Policies: Incommensurability and Uncertainty'. *Ecological Economics* 96 (2013): 132–140.

Armstrong, D. M. *A Materialist Theory of Mind*. London: Routledge, 1968.

Arnauld, Antoine, and Pierre Nicole. *Logic, Or, The Art of Thinking: Containing, besides Common Rules, Several New Observations Appropriate for Forming Judgment*. 5th ed. Cambridge, England: Cambridge University Press, 1996.

Barberis, Nicholas C. 'Thirty Years of Prospect Theory in Economics: A Review and Assessment'. *Journal of Economic Perspectives* 27, no. 1 (2013): 175–189.

Barbey A. K., F. Krueger, and J. Grafman. 'Structured Event Complexes in the Medial Prefrontal Cortex Support Counterfactual Representations for Future Planning'. *Philosophical Transactions of the Royal Society, Biological Sciences* 362 (2009): 1291–1300.

Barlow, H. B. 'The Neuron Doctrine in Perception'. In *The Cognitive Neurosciences*, edited by M. S. Gazzaniga, 415–35. Cambridge, MA: MIT Press, 1995.

Barlow-Stewart. 'Genes and Chromosomes: The Genome'. *NSW Government Health Centre for Genetics Education*. Last modified 2012. http://www.genetics.edu.au.

Bartels, Daniel M. 'Principled Moral Sentiment and the Flexibility of Moral Judgment and Decision Making'. *Cognition* 108 (2008): 381–417.

Baxter, Brian H. 'Naturalism and Environmentalism: A Reply to Hinchman'. *Environmental Values* 15 (2006): 51–68.

Bechara, A., and A. R. Damasio. 'The Somatic Marker Hypothesis: A Neural Theory of Economic Decision'. *Games and Economic Behaviour* 52 (2005): 336–372.

Bechara, A., H. Damasio, D. Tranel, and A. R. Damasio. 'Deciding Advantageously before Knowing the Advantageous Strategy'. *Science* 275, no. 5304 (1997): 1293–1295.

Bechtel, W. 'Molecules, Systems, and Behaviour: Another View of Memory Consolidation'. In Bickle, *Oxford Handbook*, 13–60.

Beheme, C. ‘A “Galilean” Science of Language’. *Journal of Linguistics* 50, no. 3 (2014): 671–704.

Bennett, M. R., and P. M. S. Hacker. *Philosophical Foundations of Neuroscience*. Blackwell, 2003.

———. ‘Reply to Professor Dennett and Professor Searle’. Accessed 15 June 2015. http://info.sjc.ox.ac.uk/scr/hacker/docs/Reply%20to%20Dennett%20and%20Searle.pdf

Benoit, R. G., S. J. Gilbert, and P. W. Burgess. ‘A Neural Mechanism Mediating the Impact of Episodic Prospection on Farsighted Decisions’. *Journal of Neuroscience* 31 (2011): 6771–79.

Bergandi, Donato, ed. *The Structural Links between Ecology, Evolution and Ethics: The Virtuous Epistemic Circle*. New York: Springer, 2013.

Berker, S. ‘The Normative Insignificance of Neuroscience’. *Philosophy and Public Affairs* 37 (2009): 293–329.

Bickle, John, ed. *The Oxford Handbook of Philosophy and Neuroscience.* New York: Oxford University Press, 2009.

Bickle, J., and A. J. Silva. ‘The Science of Research and the Search for Molecular Mechanisms of Cognitive Functions’. In Bickle, *Oxford Handbook*, 91 -129.

Blair, James R. ‘The Amygdala and Ventromedial Prefrontal Cortex in Morality and Psychopathy’. *Trends in Cognitive Science* 11 (2007): 387–392.

———. ‘Neurocognitive Models of Aggression, the Antisocial Personality Disorders, and Psychopathy’. *Journal of Neurology, Neurosurgery and Psychiatry* 71 (2001): 727–731.

Blair, James R., and L. Cipolotti. ‘Impaired Social Response Reversal: A Case of “Acquired Sociopathy”’. *Brain* 123 (2000): 1122–1141.

Botvinick, M. M. ‘Multilevel Structure in Behaviour and in the Brain: A Model of Fuster’s Hierarchy’. *Philosophical Transactions of the Royal Society, Biological Sciences* 362 (2007): 1615–1626.

Boyd, P. J., and R. Richerson. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago, IL: University of Chicago Press, 2006.

Boyer, P. *Religion Explained: The Evolutionary Origins of Religious Thought*. New York: Basic Books, 2001.

Bradie, Michael, and William Harms. ‘Evolutionary Epistemology’. In *Stanford Encyclopaedia of Philosophy*, edited by Edward N. Zalta. Stanford University, Winter 2012. http://plato.stanford.edu/archives/win2012/entries/epistemology-evolutionary/.

Brennan, Andrew, and Yeuk-Sze Lo. 'Environmental Ethics'. *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Standford University, Fall 2011. http://plato.stanford.edu/archives/fall2011/entries/ethics-environmental/.

Brigandt, Ingo, and Alan Love. 'Reductionism in Biology'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, Summer 2012. http://plato.stanford.edu/archives/sum2012/entries/reduction-biology/.

Brown, R. 'What is a Brain State?*' Philosophical Psychology* 19, no. 6 (2006): 729–42.

Buller, D. *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. Cambridge, MA: MIT Press, 2005.

Buller, T. 'What Can Neuroscience Contribute to Ethics?' *Journal of Medical Ethics* 32 (2006): 63–64.

Bureau of Ocean Energy Management, Regulation and Enforcement. Accessed 29 October 2014. http://www.boemre.gov.

Buss, David M. *Evolutionary Psychology and the New Science of the Mind.* Boston: Allyn and Bacon, 2007.

Campbell, D. T. 'How Individual and Face-to-Face Group Selection Undermine Firm Selection in Organizational Evolution'. In *Evolutionary Dynamics of Organizations*, edited by J. A. C. Baum and J. V. Singh, 23–38. New York: Oxford University Press, 1994.

Carruthers, P. *The Architecture of the Mind*. Oxford, England: Oxford University Press, 2006.

Casebeer, W. D. 'Processes and Moral Emotions'. In Sinnott-Armstrong, *Moral Psychology*, 3:19–25.

Causse, Mickaël, Frédéric Dehais, Patrice Péran, Umberto Sabatini, and Josette Pastor. 'The Effects of Emotion on Pilot Decision-Making: A Neuroergonomic Approach to Aviation Safety'. *Transportation Research Part C: Emerging Technologies* 33 (2013): 272–281.

Chalmers, D. 'What is a Neural Correlate of Consciousness?' In *Neural Correlates of Consciousness*, edited by T. Metzinger, 231-240. MIT Press, 2000.

Charney, E., and W. English. 'Candidate Genes and Political Behaviours'. *American Political Science Review* (2012): 7–10.

Churchland, Patricia S. *Brain Trust: What Neuroscience Tells Us about Morality.* Princeton, NJ: Princeton University Press, 2011.

———. 'Inference to the Best Decision'. In Bickle, *Oxford Handbook*, 419–427.

———. *Touching a Nerve: The Self as Brain*. New York: Norton and Company, 2013.

Cinelli, Marco, Stuart R. Coles, and Kerry Kirwan. ‘Analysis of the Potentials of Multi Criteria Decision Analysis Methods to Conduct Sustainability Assessment’. *Ecological Indicators* 46 (2014): 138–148.

Clune, Lipson, and Mouret. ‘The Evolutionary Origins of Modularity’. *Proceedings B: Biological Sciences*. Last modified 13 January 2013. http://rspb.royalsocietypublishing.org/content/280/1755/20122863.

Cohen, J. D. ‘The Vulcanisation of the Human Brain: A Neural Perspective on Interactions between Cognition and Emotion’. *Journal of Economic Perspectives* 19, no. 4 (2005): 3–24.

Cole, M. W, J. R. Reynolds, J. D. Power, G. Repovs, A. Anticeic, and T. S. Braver. ‘Multi-Task Connectivity Reveals Flexible Hubs for Adaptive Task Control’. *Nature Neuroscience* 16, no. 9 (2013):1348–55.

Colyvan, Mark. ‘Environmental Philosophy: Beyond Environmental Ethics’. *Arts* 29 (2007): 95–104.

Colyvan, Mark, Damian Cox, and Katie Steele. ‘Modelling the Moral Dimension of Decisions’. *NOUS* 44, no. 3 (2010): 503–529.

Conway, P., and B. Garwronski. ‘Deontological and Utilitarian Inclinations in Moral Decision Making: A Process Dissociation Approach’. *Journal of Personality and Social Psychology* 104, no. 2 (2013): 216–235.

Cosmides, Leda. ‘The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task’. *Cognition* 31 (1989): 187–276.

Cosmides, Leda, and John Tooby. ‘Can a General Deontic Logic Capture the Facts of Human Moral Reasoning? How the Mind Interprets Social Exchange Rules and Detects Cheaters’. In Sinnott-Armstrong, *Moral Psychology* 1:53–120.

———. ‘Cognitive Adaptations for Social Exchange’. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, edited by J. H. Barkow, L. Cosmides, and J. Tooby, 163–228. Oxford: Oxford University Press, 1992.

Craver, C, F., and K. S. Robins. ‘Biological Clocks: Explaining with Models of Mechanisms’. In Bickle, *Oxford Handbook*, 41–67.

Crick, F. C., and Koch, C. ‘What is the Function of the Claustrum?’ *Philosophical Transactions of the Royal Society, Biological Sciences* 360 (2005): 1271–1279.

Crockett, M. J. ‘Models of Morality’. *Trends in Cognitive Science* 17 (2013): 363–366.

Cushman, F. ‘The Neural Basis of Morality: Not Just Where, But When’. *Brain* 137 (2014): 974–980.

Damasio, A. R. *Descartes’ Error: Emotion, Reason and the Human Brain*. New York, NY: Putnam, 1994.

Darwin, Charles. *The Descent of Man, and Selections in Relation to Sex*, 1871.

———. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. 1850. Reprint, New York: Signet Classics, 2003.

———. *The Variation of Animals and Plants under Domestication*. London: John Murray, 1868.

Dawkins, Richard. *The Selfish Gene*. New York, NY: Oxford University Press, 1976.

DeLisi, M., K. M. Beaver, M. G. Vaughn, and J. P. Wright. ‘All in the Family’. *Criminal Justice and Behaviour* 36, no. 11 (2009): 1187–1197.

De Neys, W., and T. Glumicic. ‘Conflict Monitoring in Dual Process Theories of Reasoning’. *Cognition* 106 (2008): 1248–1299.

Dennett, Daniel. ‘Three Kinds of Intentional Psychology (IP)’. In *Philosophy of Mind: A Guide and Anthology*, edited by J. Heil. Oxford: Clarendon Press, 2004.

de Waal, Frans, ed. *Primates & Philosophers: How Morality Evolved*. Princeton, NJ: Princeton University Press, 2006.

Dewey, J. *Human Nature and Conduct: An Introduction to Social Psychology*. In Flanagan, O., H. Sarkissian, and D. Wong, ‘Naturalizing Ethics’. In Sinnott-Armstrong, *Moral Psychology*, 1:1–27.

di Pellegrino, G., L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti. ‘Understanding Motor Events: A Neurophysiological Study**’.** *Experimental Brain Research* 91 (1992): 176–180.

Dietrich, Michael R. ‘A Just-So Story for Symbolic Thought? Comment on Tse’. In Sinnott-Armstrong, *Moral Psychology*, 1:299–302.

Doolittle, W. F. ‘The Practice of Classification and the Theory of Evolution, and What the Demise of Charles Darwin’s Tree of Life Hypothesis Means for Both of Them’. *Philosophical Transactions of the Royal Society B*, 264 (2009): 2221.

Downes, Stephen M. ‘Evolutionary Psychology’. In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, Fall 2010. http://plato.stanford.edu/archives/fall2010/entries/evolutionary-psychology/.

Duncan, J. ‘An Adaptive Coding Model of Neural Function in Prefrontal Cortex’. *Nature Reviews Neuroscience* 2, no. 11 (2001): 820–9.

Dyck, B. A., K. J. Skoblenick, J. M. Castellano, K. Ki, N. Thomas, and R. K. Mishra. ‘Behavioural Abnormalities in Synapsin 11 Knockout Mice Implicate a Causal Factor in Schizophrenia’. *Synapse* 63, no. 8: 662–672.

Eastman, N., and B. Starling. ‘Mental Disorder Ethics: Theory and Empirical Investigation’. *Journal of Medical Ethics* 32 (2005): 79–83.

Edelman, Geral M. ‘Neural Darwinism: Selection and Reentrant Signaling in Higher Brain Function’. *Neuron* 10 (1993): 115–125.

Edwards, W., R. F. Miles, and D. Winterfeldt, eds. *Advances in Decision Analysis: From Foundations to Applications*. New York: Cambridge University Press, 2007.

Erlich, Paul R. *Human Natures: Genes, Cultures, and the Human Prospect.* Washington DC: Island Press, 2000.

Ehrlich, Paul R., and Anne H. Ehrlich. ‘Influencing Policy’. In Van de Veer and Pierce, *Environmental Ethics and Policy Book*, 621–629.

Ellis, George. ‘View from the Top’. *New Scientist*, 17 August 2013.

Eslinger, P. J., and A. R. Damasio. ‘Severe Disturbance of Higher Cognition after Bilateral Frontal Lobe Ablation: Patient EVE’. *Neurology* 35 (1985): 1731–1741.

Feigl, H. *The ‘Mental’ and the ‘Physical’: The Essay and a Postscript*. Minneapolis, MN: University of Minnesota Press, 1967.

Firat, R., and C. M. McPherson. ‘Toward an Integrated Science of Morality’. In *Handbook of the Sociology of Morality*, edited by S. Hitlin and S. Vaisey, 361–384. New York: Springer, 2010.

Firth, C. D. *Making Up the Mind: How the Brain Creates Our Mental World*. Oxford: Blackwell, 2007.

Fisher, R. A. *The Genetical Theory of Natural Selection*. Oxford, England: Clarendon Press, 1930.

Fitzpatrick, William. ‘Morality and Evolutionary Biology’. In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, Summer 2012. http://plato.stanford.edu/archives/sum2012/entries/morality-biology.

Flanagan, Owen. *Varieties of Moral Personality.* Cambridge, MA: Harvard University Press, 1991.

Flanagan, O., H. Sarkissian, and D. Wong. ‘Naturalizing Ethics’. In Sinnott-Armstrong, *Moral Psychology*, 1:1–27.

Flint, J., R. Greenspan, and K. Kendler. *How Genes Influence Behaviour*. New York, NY: Oxford University Press, 2010.

Fodor, Jerry A. *The Modularity of Mind*. Cambridge, MA: MIT Press, 1983.

Foot, P. 'The Problem of Abortion and the Doctrine of Double Effect'. *Oxford Review* 5 (1967): 5–15.

Friedman, Milton. *Essays in Positive Economics.* Chicago: University of Chicago Press, 1953.

Fuster, J. M. 'Upper Processing Stages of the Perception: Action Cycle'. *Trends in Cognitive Science* 8 (2004): 143–145.

Gaissmaier, W., and G. Gigerenzer. 'Heuristic Decision Making'. *Annual Review of Psychology* 62 (2011): 451–482.

Gazzaniga, Michael S. *The Ethical Brain: The Science of Our Moral Dilemmas*. New York: Harper-Perennial, 2006.

Geary, D. C., and K. J. Huffman. 'Brain and Cognitive Evolution: Forms of Modularity and Functions of Mind'. *Psychological Bulletin* 128 (2002): 667–698

Ghiselin, M. T. *The Economy of Nature and the Evolution of Sex*. Berkeley, CA: University of California Press, 1974.

Gigerenzer, Gerd. 'Moral Intuition = Fast and Frugal Heuristics?' In Sinnott-Armstrong, *Moral Psychology,* 2:1–26.

Gillett, G. 'Cyborgs: Human Machine Hybrids and Moral Identity'. *Journal of Medical Ethics* 32 (2005): 79–83.

Glimcher, Paul. 'Choice: Towards a Standard Back-Pocket Model'. In *Neuroeconomics: Decision Making and the Brain,* edited by P. Glimcher, C. Camerer, E. Fehr, and R. Poldrack, 503–521. New York: Academic Press, 2000. http://www.decisionsrus.com/documents/choice_towards-a-standard-back-pocket-model.pdf.

Goldstein, D. G., and G. Gigerenzer. 'Models of Ecological Rationality: The Recognition Heuristic'. *Psychological Review* 109 (2002): 75–90.

Goswami, U. 'Neuroscience and Education: From Research to Practice?' *Nature Reviews Neuroscience* 7, no. 5 (2006): 406–413.

Gould, Stephen Jay. *An Evolutionary Perspective on Strengths, Fallacies, and Confusions in the Concept of Native Plants*. Harvard Arnoldia, 1998.

———. *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press, 2002.

———. *Wonderful Life: The Burgess Shale and the Nature of History*. New York, NY: W. W. Norton, 1990.

Grabenhorst, F., and E. T. Rolls. 'Value, Pleasure and Choice in the Ventral Prefrontal Cortex'. *Trends in Cognitive Science* 15 (2011): 56–67.

Greene, Joshua. 'Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics'. *Ethics* 124, no. 4 (2014): 695–726.

———. 'The Cognitive Neuroscience of Moral Judgment'. In *The Cognitive Neurosciences*, 5th ed., edited by M. S. Gazzaniga, 987–1002. Cambridge: MIT Press, 2009.

———. *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York, NY: The Penguin Press, 2013.

———. 'Notes on "The Normative Insignificance of Neuroscience" by Selim Berker'. Accessed 1 March 2014. http://www.wjh.harvard.edu/~jgreene/GreeneWJH/Greene-Notes-on-Berker-Nov10.pdf.

———. 'The Secret Joke of Kant's Soul'. In Sinnott-Armstrong, *Moral Psychology*, 35–80.

Greene, J. D., S. A. Morelli, K. Lowenberg, L. E. Nystrom, and J. D. Cohen. 'Cognitive Load Selectively Interferes with Utilitarian Moral Judgment'. *Cognition* 107 (2008): 1144–1154.

Greene, J. D., L. E. Nystrom, A. D. Engell, J. M. Darley, and J. D. Cohen. 'The Neural Bases of Cognitive Conflict and Control in Moral Judgment'. *Neuron* 44 (2004): 389–400.

Grinspoon, David. 'Welcome to Terra Sapiens', *Aeon (Online Magazine)*, 20 December 2016, https://aeon.co/essays/enter-the-sapiezoic-a-new-aeon-of-self-aware-global-change

Grüne-Yanoff, Till. 'Bounded Rationality'. *Philosophy Compass* 2/3 (2007): 534–563.

Haidt, Jonathan. 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment'. *Psychological Review* 108, no. 4 (2001): 814–34.

Haldane, J. B. S. 'A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation'. *Proceedings of the Cambridge Philosophical Society* 23 (1927): 838.

Hansson, Sven Ove. 'The Harmful Influence of Decision Theory on Ethics'. *Ethical Theory and Moral Practice* 13 (2010): 585–593.

Harle, K. M., L. J. Chang, M. van't Wout, and A. G. Sanfey. 'The Neural Mechanisms of Affect Infusion in Social Economic Decision-Making: A Mediating Role of the Anterior Insula'. *NeuroImage* 61 (2012):32–4.

Harrison, J. 'Ethical Objectivism'. In *The Encyclopaedia of Philosophy*: *Volumes 3–4*, edited by P. Edwards, 71–75. New York: Macmillian, 1967.

Henrich, J., R. Boyd, and P. J. Richerson. 'Five Misunderstandings about Cultural Evolution'. *Human Nature* 19 (2008): 119–137.

Haug, M. C. 'Introduction: Debates about Methods: From Linguistic Philosophy to Philosophical Naturalism'. In *Philosophical Methodology: The Armchair or the Laboratory?* edited by M. C. Haug, 1–26. New York: Routledge, 2014.

Hauser, Larry. 'Behaviorism'. *Internet Encyclopedia of Philosophy*. Last modified 19 July 2005. http://www.iep.utm.edu/behavior/#SSH1a.ii.

Hauser, M. D. 'The Liver and the Moral Organ'. *Social Cognitive and Affective Neuroscience* 1 (2006): 214–220.

———. *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. Harper Collins, 2006.

Heyser, C., and A. Chemero. 'Methodology and Reduction in the Behavioural Neurosciences: Object Exploration as a Case Study'. In Bickle, *Oxford Handbook*, 68–90.

Hinchman, Lewis. 'Is Environmentalism a Humanism?' *Environmental Values* 13 (2004): 3–29.

Howard, Ronald A. 'The Foundations of Decision Analysis Revisited'. In Edwards, Miles, and Winterfeldt, *Advances in Decision Analysis*, 32–57.

Hull, David L. 'Units of Evolution: A Metaphysical Essay'. In *The Philosophy of Evolution*, edited by U. J. Jensen and R. Harré, 23–44. Brighton, England: The Harvester Press, 1981.

Hume, David. *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge and P. H. Nidditch. Oxford, UK: Oxford University Press, 1739/1978.

Haushofer J, Fehr E, 'You shouldn't have: your brain on others' crimes'. *Neuron* 60, (2008): 738-740.

Hynes, C. A. 'Morality, Inhibition, and Propositional Content'. In Sinnott-Armstrong, *Moral Psychology*, 3:25–30.

Illes, J., and E. Racine. 'Imaging or Imagining? A Neuroethics Challenge Informed by Genetics'. *American Journal of Bioethics* 5 (2005): 1–14.

ISO. ISO31000/31010: ‘Risk Management & Risk Assessment Techniques’, http://www.iso.org/iso/home/standards/iso31000.htm Accessed 24/6/15

Jablonka, E., and M. Lamb. *Evolution in Four Dimensions: Genetic, Epigentic, Behavioural, and Symbolic Variation in the History of Life*. Cambridge MA: MIT Press, 2005.

James, Scott M. *An Introduction to Evolutionary Ethics*. Hoboken, NJ: Wiley, 2010. http://UNE.eblib.com.au/patron/FullRecord.aspx?p=624762.

James, W. *The Principles of Psychology*. New York, NY: Henry Holt, 1890.

Johnson, J. G., and M. Raab . ‘Take The First: Option Generation and Resulting Choices’. *Organizational Behavior and Human Decision Processes* 91 (2003): 215–229.

Joyce, James M. *The Foundations of Causal Decision Theory*. Cambridge University Press, 1999.

Joyce, R. *The Evolution of Morality*. Cambridge: Massachusetts Institute of Technology Press, 2006.

———. ‘What Neuroscience Can (and Cannot) Contribute to Metaethics’. In Sinnott-Armstrong, *Moral Psychology*, 3:371–394.

Kable, J., and P. Glimcher. ‘The Neurobiology of Decision: Consensus and Controversy’. *Neuron* 63 (2009): 733–745.

Kant, Immanuel. *Grounding for the Metaphysics of Morals* (3rd ed., first published 1785). Translated by James W. Ellington. Hackett, 1993.

Kendal, J. R., and K. N. Laland. ‘Mathematical Models for Memetics’. *Journal of Memetics* 4 (2000). http://cfpm.org/jom-emit/2000/vol4/kendal_jr&laland_kn.html.

Kelman, S. ‘Cost Benefit Analysis: An Ethical Critique’. In Van de Veer and Pierce, *Environmental Ethics and Policy Book*, 329–335.

Kennett, J., and C. Fine. ‘Internalism and the Evidence from Psychopaths and “Acquired Sociopaths”’. In Sinnott-Armstrong, *Moral Psychology*, 3:173–190.

Kilner, J. M., and R. N. Lemon. ‘What We Know Currently about Mirror Neurons’. *Current Biology* 23, no. 23 (2013). http://www.cell.com/current-biology/abstract/S0960-9822(13)01326-2#MainText).

Kimura, K., S. Izawa, N. Sugaya, N. Ogawa, and K. C. Yamada. ‘The Biological Effects of Acute Psychosocial Stress on Delay Discounting’. *Psychoneuroendocrinology* 38, no. 10 (2013): 2300–8.

King, J. A., J. R. Blair, D. G. Mitchell, R. J. Dolan, and N. Burgess. 'Doing the Right Thing: A Common Neural Circuit for Appropriate Violent or Compassionate Behaviour'. *Neuroimage* 30 (2006): 1069–1076.

Kirkman, Robert. 'Darwinian Humanism: A Proposal for Environmental Philosophy'. *Environmental Values* 16 (2007): 3–21.

Kitcher, Phillip. 'Ethics and Evolution: How to Here from There'. In de Waal, *Primates & Philosophers*, 120 –140.

———. *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. Cambridge, MA: Massachusetts Institute of Technology Press, 1985. In James, *An Introduction to Evolutionary Ethics*.

Kleffner D. A., and V. S. Ramachandran. 'On the Perception of Shape from Shading'. *Perception and Psychophysics* 52, no. 1 (1992): 18–36.

Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. A. Cushman, M. Hauser, and A. Damasio. 'Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements'. *Nature* 446 (2007): 908–911.

Kohlberg, Lawrence. *Essays in Moral Development.* Vol I, *The Psychology of Moral Development.* New York, NY: Harper and Row, 1984.

———. 'Moral Development and Identification'. In *Child Psychology: The Sixty-Second Yearbook of the National Society for the Study of Education* (Part 1), edited by H. Stevenson, 277–332. Chicago, IL: University of Chicago Press, 1963.

Korsgaard, C. 'Morality and the Distinctiveness of Human Action'. In de Waal, *Primates & Philosophers*, 98-119.

Laland, K. N., and G. R. Brown. *Sense and Nonsense*. Oxford: Oxford University Press, 2002.

Landreth, Anthony. 'The Emerging Theory of Motivation'. In Bickle, *Oxford Handbook,* 381–418.

Larrere, Catherine. 'Two Philosophies of the Environmental Crisis'. In Bergandi, *Structural Links*, 142–149.

Laughlin, C., and J. Loubser, 'Neurognosis, the Development of Neural Models, and the Study of the Ancient Mind'. *Time and Mind: The Journal of Archaeology, Consciousness and Culture* 3, no. 2 (2010): 146.

Lennox, James. 'Darwinism'. In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, Fall 2010. http://plato.stanford.edu/archives/fall2010/entries/darwinism/.

Leslie, A. M. ‘Pretence and Representation: The Origins of “Theory of Mind”’. *Psychological Review* 94 (1987): 412–426.

Levitin, Daniel J. *This Is Your Brain On Music*. New York, NY: Penguin, 2007.

Lewens, Tim. ‘Cultural Evolution’. In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, Spring 2013. http://plato.stanford.edu/archives/spr2013/entries/evolution-cultural/.

———. ‘*Cultural Evolution: Integration and Skepticism*’. In *The Oxford Handbook of Philosophy and Social Science*, edited by H. Kingcaid, 458–480. New York: Oxford University Press, 2012.

Lighthall N. R., M Sakaki, S. Vasunilashorn, L. Nga, and S. Somayajula. ‘Gender Differences in Reward Replated Decision Processing under Stress’. *Social Cognition Affect Neuroscience* 7 (2012):476–84.

Lloyd, Elisabeth. ‘Units and Levels of Selection’. In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, Winter 2012. http://plato.stanford.edu/archives/win2012/entries/selection-units/.

Locke, John. *Essay Concerning Human Understanding*. 1689. Reprint, New York: Dover, 1959.

Looren de Jong, H. ‘Evolutionary Psychology and Morality. Review Essay’. *Ethical Theory and Moral Practice* 14, no.1 (2011): 117–125.

Louie, Khaw, and Glimcher. ‘Normalization is a General Neural Mechanism for Context-Dependent Decision Making’. *Psychological and Cognitive Sciences* 110, no.15 (2013): 6139–6144.

Lozano, Andres. ‘Parkinson’s, Depression and the Switch that Might Turn Them Off’. January 2013 at TEDxCaltech.

Lynch, Tony. ‘Understanding What Sustainable Development is and What It is Not’. *Philosopher.io* (blog), 6 October 2015, http://philosopher.io/Understanding-What-Sustainable-Development-Is-And-What-It-Is-Not.

———. ‘Winnicott’s World: The Mystical Foundations of Reality’. *Transcendent Philosophy* 4, no. 4 (2003).

Lynch, Tony, and David Wells. *The Political Ecologist*. Ashgate, 2000.

McDermott, R., D. Tingley, and T. J. Cowden. ‘Monoamine Oxidase A Gene (MAOA) Predicts Behavioural Aggression Following Provocation’. *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 7 (2009): 2118–2123.

Magdoff, F., and F. Bellamy. ‘What Every Environmentalist Needs to Know about Capitalism’. In Pojmon and Pojmon, *Environmental Ethics*, 69–709.

Mangagi, Shunsuke. *Technological Change and Environmental Policy: A Study of Depletion in the Oil and Gas Industry*. Cheltenham, UK: Edward Elgar, 2007.

Maynard Smith, J. ‘Natural Selection of Culture?’ *New York Review of Books*, 6 November 1986, 33.

Mayr, E. *Toward a New Philosophy of Biology*. Cambridge, MA: Harvard University Press, 1988.

McGinn, C. ‘Can We Solve the Mind–Body Problem?’ In *The Mind–Body Problem: A Guide to the Current Debate*, edited by R. Warner and T. Szubka, 349-346. Oxford, England: Blackwell, 1994.

Mesoudi, A. *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago, IL: University of Chicago Press, 2011.

Mesoudi, A., and M. J. O’Brien. ‘The Cultural Transmission of Great Basin Projectile Point Technology I: An Experimental Simulation’. *American Antiquity* 73, no. 1 (2008): 3–28. http://amesoudi.googlepages.com/Mesoudi_OBrien_AmAnt_2008a.pdf.

Mikhail, John. *Elements of Moral Cognition: Rawls’ Linguistic Analogy and the Cognitive Science of Moral and Legal Judgement*. New York, NY: Cambridge University Press, 2011.

Miller, E. K., and J. D. Cohen. ‘An Integrative Theory of Prefrontal Cortex Function’. *Annual Review of Neuroscience* 24 (2001): 167–202. http://search.proquest.com.ezproxy.une.edu.au/docview/198920013?accountid=17227.

Minteer, Ben A, *Refounding Environmental Ethics: Pragmatism, Principle and Practice*. Philadelphia: Temple University Press, 2012.

Moll, Jorge. ‘The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions’. *Journal of Neuroscience* 22, no. 7 (2002): 2730–2736.

Moll, J., R. Zahn, R. de Oliveira-Souza, F. Krueger, and J. Grafman. ‘The Neural Basis of Human Moral Cognition’. *Nature Reviews Neuroscience* 6 (2005): 799–809.

Moll, J., R. Oliveira-Souza, R. Zhan, and J. Grafman. ‘The Cognitive Neuroscience of Moral Emotions’. In Sinnott-Armstrong, *Moral Psychology*, 1–18.

National Institute of Neurological Disorders and Stroke (USA). http://www.ninds.nih.gov.

Nicolis, Gregoire, and Catherine Rouvas-Nicolis. ‘Complex Systems’. *Scholarpedia* 2, no. 11 (2007): 1473. www.scholarpedia.org/article/complex_systems.

Norton, Bryan G. ‘Facts, Values, and Analogies: A Darwinian Approach to Environmental Choice’. In Bergandi, *Structural Links*.

———. ‘Leopold, Hadley, and Darwin: Darwinian Epistemology, Truth, and Right’. *Contemporary Pragmatism* 10, no. 1 (2013): 1–28.

———. *Sustainability: A Philosophy of Adaptive System Management*. Chicago, IL: University of Chicago Press, 2005.

Nguyen, Y., and C. N. Noussair. ‘Risk Aversion and Emotions’. *Pacific Economic Review* 1, no. 3 (2014): 296–312.

Ophuls, W. *Plato’s Revenge: Politics in the Age of Ecology*. Cambridge, MA: MIT Press, 2011.

Ortmann, A., G. Gigerenzer, B. Borges, and D. G. Goldstein. ‘The Recognition Heuristic: A Fast and Frugal Way to Investment Choice? In *The Handbook for Experimental Economics Results*, edited by C. R. Plott and V. L. Smith, 1(28): 993–1003. Amsterdam: Elsevier, 2008.

Ostrom, Elinor. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990.

Oyama, S. *The Ontogeny of Information: Developmental Systems and Evolution*. Cambridge, England, 1985.

Oyama S., P. E. Griffiths, and R. D. Gray, eds. *Cycles of Contingency: Developmental Systems and Evolution.* Cambridge: Massachusetts Institute of Technology Press, 2001.

Padoa-Schioppa, C., and J. A. Assad. ‘Neurons in the Orbitofrontal Cortex Encode Economic Value’. *Nature* 441 (2006): 223–226.

Palmer, C. ‘An Overview of Environmental Ethics’. In Pojmon and Pojmon, *Environmental Ethics*, 10–36.

Pascual, L., D. Gallardo-Pujol, and P. Rodrigues. ‘How Does Morality Work in the Brain? A Functional and Structural Perspective of Moral Behaviour’. *Frontiers in Integrative Neuroscience* 7, no. 65 (2013).

Paxton, J. M., L. Ungar, and J. D. Greene. ‘Reflection and Reasoning in Moral Judgment’. *Cognitive Science* 36 (2011): 167–177.

Pellikaan, Huib, and Robert J. Van der Veen. *Environmental Dilemmas and Policy Design*. Cambridge, UK: Cambridge University Press, 2002.

Perry. ‘Neurons: A Curious Collection of Shapes and Sizes’. *BrainFacts.org*. Last modified 22 August 2012. http://www.brainfacts.org/brain-basics/neuroanatomy/articles/2012/neurons-a-curious-collection-of-shapes-and-sizes/.

Peterson, Martin. *An Introduction to Decision Theory*. New York: Cambridge University Press, 2009.

Phelps, E. A. 'The Study of Emotion in Neuroeconomics'. In *Neuroeconomics: Decision Making and the Brain*, edited by P. Glimcher, C. Camerer, E. Fehr, and R. Poldrack, 229–243. New York, NY: Academic Press, 2000. http://www.hse.ru/data/2011/12/09/1259101786/Neuroeconomics.pdf

Phelps, E. A, K. M. Lempert, and P. Sokol-Hessner. 'Emotion and Decision Making: Multiple Modulatory Neural Circuits'. *Annual Review Neuroscience* 37 (2014): 263–87.

Pinker, Steven. *How the Mind Works.* New York: W. W. Norton & Company, 1997.

Place, U. T. 'Is Consciousness a Brain Process?' In *Identifying the Mind: Selected Papers of U. T. Place*, edited by G. Graham and E. Valentine, 70-82. Oxford University Press, 2004.

Plotkin, Henry. *Darwin Machines and the Nature of Knowledge*. Harvard University Press, 1994.

Pojmon, L., and P. Pojmon, eds. *Environmental Ethics: Readings in Theory and Application*, 6th ed. Wadsworth, 2012.

Polasky, Stephen, Stephen R. Carpenter, Carl Folke, and Bonnie Keeler. 'Decision-Making under Great Uncertainty: Environmental Management in an Era of Global Change'. *Trends in Ecology and Evolution* 26, no. 8 (2011).

Polkinghorne, J. 'Reductionism'. *The Interdisciplinary Encyclopedia of Religion and Science*, 2002. http://inters.org/reductionism.

Porcelli, A. J., and M. R. Delgado. 'Acute Stress Modulates Risk Taking in Financial Decision Making'. *Psychological Science* 20 (2009): 278–83.

Prinz, Jesse. 'Is Morality Innate?' In Sinnott-Armstrong, *Moral Psychology,* 1:367–406.

———. 'Reply to Dwyer and Tiberius'. In Sinnott-Armstrong, *Moral Psychology,* 1:427–440.

Racine, E., and J. Illes. *Is Neuroethics the Heir of the Ethics of Genomics?* Calgary, Alberta: Canadian Bioethics Society, 2004.

Railton, P. 'Naturalism Relativized'. In Sinnott-Armstrong, *Moral Psychology,* 1:37–44..

Raine, A. 'From Genes to Brain to Antisocial Behaviour'. *Current Directions in Psychological Science* 17, no. 5 (2008): 323–328.

Ramachandran, V. S. *The Tell-Tale Brain: Unlocking the Mystery of Human Nature*. London: Windmill Books, 2012.

Rees, William E. ‘Sustainable Development: Economic Myths and Global Realities’. In Pojmon and Pojmon, *Environmental Ethics*, 661–8.

Robbins, Philip. ‘Modularity of Mind’. In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, Summer 2010. http://plato.stanford.edu/archives/sum2010/entries/modularity-mind/.

Rosenberg, A. ‘Lessons for Cognitive Science from Nuerogenomics’. In Bickle, *Oxford Handbook*, 143–165.

Rosenberg, A. ‘Will Genomics Do More for Metaphysics than Locke?’ In *Evolutionary Ethics and Contemporary Biology*, edited by G. Boniolo and G. De Anna, 178–98. 2006.

Roskies, A. L. ‘What’s “NEU” in Neuroethics?’ In Bickle, *Oxford Handbook*, 454–468.

Routley, Richard. ‘Is There a Need for a New, an Environmental, Ethic? Proceedings of the XVth World Congress of Philosophy’ (1973) 1: 205-210.

Rowley, Hazel V., Gregory M. Peters, Sven Lundie, and Stephen J. Moore. ‘Aggregating Sustainability Indicators: Beyond the Weighted Sum’. *Journal of Environmental Management* 111 (2012): 24–33.

Royzman, Edward B., and J. Baron. ‘The Preference for Indirect Harm’. *Social Justice Research* 15 (2002): 165–84.

Rozzi, R. ‘The Reciprocal Links between Evolutionary Ecological Sciences and Environmental Ethics’. *Bioscience* 49, no. 11 (1999): 911–921.

Ryle, Gilbert. *The Concept of Mind*, 1949.

Sarkar, Sahotra. *Environmental Philosophy: From Theory to Practice*. John Wiley & Sons, 2012.

———. ‘The Inevitability of Normative Analysis’. *Behavioural and Brain Sciences* 37, no. 4 (2014): 436.

———. ‘Multiple Criteria and Trade-Offs in Environmental Ethics: Comment on “Ethics of Species Research and Preservation” by Rob Irvine’. *Bioethical Inquiry* 10 (2013): 533–537.

Sandberg, A., and N. Bostrom. ‘Whole Brain Emulation: A Roadmap’. Technical Report #2008‐3. Oxford, England: Future of Humanity Institute, Oxford University, 2008. www.fhi.ox.ac.uk/reports/2008‐3.pdf.

Samejima, K., Y. Ueda, K. Doya, and M. Kimura. ‘Representation of Action-Specific Reward Values in the Striatum’. *Science* 310 (2005): 1337–1340.

Saniots, A. ‘Making Connectivities: Neuroanthropology and Ecological Ethics’. *Neuroquantology* 8, no. 2 (2010): 200–205.

Savulescu, J., W. Sinnott-Armstrong, N. Levy, and B. Fulford. **‘**Neuroscience of Morality**’.** *Oxford Centre for Neuroethics***.** Accessed 1 March 2014. http://www.neuroethics.ox.ac.uk/research/area_4.

Schirmann, F. ‘Invoking the Brain in Studying Morality: A Theoretical and Historical Perspective on the Neuroscience of Morality’. *Theory & Psychology* 23, no. 4 (2013): 290–304.

Schofield, Michael. *Plato: Political Philosophy*. New York, NY: Oxford University Press, 2006.

Schultz, W. ‘Reward Signalling by Dopamine Neurons’. *Neuroscientist* 7, no. 4 (2001): 293–302.

Schultz, W., P. Dayan, and P. R. Montague. ‘A Neural Substrate of Prediction and Reward’. *Science* 275, no. 5306 (1997): 1593–1599.

Schwabe L, and O. T. Wolf. ‘Stress Prompts Habit Behaviour in Humans’. *Journal of Neuroscience* 39, no. 22 (2009): 7191–98.

Searle, John R. ‘Consciousness’. In *A Cognitive Science Dialogue: Consciousness East and West*. 2005.

———. *The Mystery of Consciousness*. 1997.

Shallice, Tim. *From Neuropsychology to Mental Structure.* Cambridge University Press, 1988.

Simon, Herbert A. ‘Bounded Rationality in Social Science: Today and Tomorrow’. *Mind and Society* 1, no. 1 (2000): 25–39.

Singer, Peter. *Practical Ethics*. Cambridge: Cambridge University Press, 1993.

Singer, Wolf. ‘Neuronal Synchronization: A Solution to the Binding Problem?’ In *The Mind-Brain Continuum*, edited by R. Llinas and P. Churchland, 101–130. The MIT Press, 1996.

Sinnott-Armstrong, W., ed. *Moral Psychology*. Vol. 1, *The Evolution of Morality: Adaptions and Innateness*. Cambridge: Massachusetts Institute of Technology Press, 2008.

———, ed. *Moral Psychology*. Vol. 2, *The Cognitive Science of Morality: Intuition and Diversity*. Cambridge: Massachusetts Institute of Technology Press, 2008.

———, ed. *Moral Psychology*. Vol. 3, *The Neuroscience of Morality: Emotion, Brain Disorders and Development*. Cambridge: Massachusetts Institute of Technology Press, 2008.

Slaby, Jan, 'Crossing the Is/Ought Boundary', *The Hedgehog Review, 2013 46-54.* http://janslaby.com/downloads/slaby_moralityhedgehohreview_2013.pdf

Smart, J. J. C. 'Sensations and Brain Processes'. In *The Nature of Mind*, edited by D. M. Rosenthal, 169–176. Oxford, England: Oxford University Press, 1991.

Smit, H., and P. Hacker. 'Seven Misconceptions about the Merelogical Fallacy: A Compilation for the Perplexed'. *Erkenntnis* 79, no. 5 (2014): 1077–1097.

Sober, E. 'Models of Cultural Evolution'. In *Trees of Life: Essays in the Philosophy of Biology*, edited by P. Griffiths, 477–492. Kluwer, 1991.

Sober, E., and D. S. Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behaviour*. Cambridge, MA: Harvard University Press, 1999.

Spencer, Herbert. *Data of Ethics*. 1883. Reprint, Belle Fourche, SD: Kessinger, 1998.

Sperber, Dan, Nicolas Baumard, and Jean-Batiste André. 'A Mutualistic Approach to Morality: The Evolution of Fairness by Partner Choice'. *Behavioural and Brain Sciences* 36 (2013): 59–112.

Sripada, Chandra. 'Nativism and Moral Psychology: The Models of the Innate Structure that Shapes the Contents of Moral Norms'. In Sinnott-Armstrong, *Moral Psychology*, 1:319–345.

Sripada, Chandra, and S. Stich. 'A Framework for the Psychology of Norms'. In *The Innate Mind.* Vol. 2, *Culture and Cognition*, edited by P. Carruthers, S. Laurence, and S. Stich, 280–301. New York: Oxford University Press, 2006.

Sustain, C. R. 'Moral Heuristics'. *Behaviour Brain Science* 28 (2005): 531–573.

Suter, R. S., and R. Hertwig. 'Time and Moral Judgment'. *Cognition* 119 (2011): 454–458

Strohminger, N. et al. 'Divergent Effects of Different Positive Emotions on Moral Judgment'. Cognition 119, no. 2 (2011.): 295–300.

Taber-Thomas, B., E. Asp, M. Koenigs, M. Sutterer, S. Anderson, and D. Tranel. 'Arrested Development: Early Prefrontal Lesions Impair the Maturation of Moral Judgment'. *Brain* 137 (2014): 1254–1261.

Tainter, J. A., and T. W. Patzek. *Drilling Down: The Gulf Oil Debacle and Our Energy Dilemma*. New York: Springer, 2012.

Tanner, Carmen. 'To Act or Not to Act: Nonconsequentialism in Environmental Decision Making'. *Ethics and Behaviour* 19, no. 6 (2009): 479–495.

Tegmark, Max. *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*. Vintage Books, 2014.

Thiele, Paul L. *Indra's Net and the Midas Touch: Living Sustainably in a Connected World*. Cambridge, MA: MIT Press, 2011.

Thomson, J. J. 'The Trolley Problem'. *Yale Law Journal* 94 (1985): 1395–1415.

Tiberius, Valerie. 'The Nativism Debate and Moral Philosophy: Comments on Prinz'. In Sinnott-Armstrong, *Moral Psychology*, 1:419–426.

Tisak, M., and E. Turiel. 'Children's Conceptions of Moral and Prudential Rules'. *Child Development* 55 (1984): 1030–1039.

Tooby, J., and L. Cosmides. 'Conceptual Foundations of Evolutionary Psychology'. In *The Handbook of Evolutionary Psychology*, edited by D. Buss, Hoboken, NJ: Wiley, 2005.

Traer, Robert. *Doing Environmental Ethics*. West View Press, 2011.

Tse, Peter Ulric. 'Symbolic Thought and the Evolution of Human Morality'. In Sinnott-Armstrong, *Moral Psychology*, 1:269–298.

Uttal, William R. *The New Phrenology*, MIT Press, 2001.

Valiant, Leslie. *Probably Proximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. New York: Basic Books, 2013.

Van de Veer, D., and C. Pierce, eds. *The Environmental Ethics & Policy Book*. Belmont, CA: Wadsworth, 1994.

Van Lier, J., R. Revlin, and W. De Neys. 'Detecting Cheaters without Thinking: Testing the Automaticity of the Cheater Detection Module'. *PLoS ONE* 8, no. 1 (2010): e53827. doi:10.1371/journal.pone.0053827.

Viren, S. *Evolutionary Psychology*. West Sussex: John Wiley & Sons, 2011.

Wallace, K. 'Morality and the Capacity for Symbolic Cognition: Comment of Tse'. In Sinnott-Armstrong, *Moral Psychology*, 1:303–315.

Warneken, F. and Tomasello, M. 'Helping and Cooperation at 14 Months of Age'. *Infancy* 11 no. 3 (2007): 271–294.

Wason P. C. 'Reasoning'. In *New Horizons in Psychology I*, edited by B. M. Foss, 106–137. Harmondsworth, England: Penguin, 1966.

West, S. A., C. El Mouden, and A. Gardner. '16 Common Misconceptions about the Evolution of Cooperation in Humans'. *Evolution and Human Behaviour* (in press). http://qqq.zoo.ox.ac.uk/group/west/pdf/west_etal.pdf.

Williams, Bernard. *Ethics and the Limits of Philosophy*. London: Routledge Classics, 2006 (first published 1985).

Williams, George C. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton, NJ: Princeton University Press, 1966.

———. 'Reply to Comments on "Huxley's Evolution and Ethics in Sociobiological Perspective"'. *Zygon* 23 (1988): 655–64.

Wilson, David Sloan. *Does Altruism Exist? Culture, Genes and the Welfare of Others.* New Haven, NJ: Yale University Press, 2015.

Wilson, E. O. *Consilience: The Unity of Knowledge.* Random House, New York. 1998.

Wood, J. N., and J. Grafman. 'Human Prefrontal Cortex: Processing and Representational Perspectives'. *Nature Reviews Neuroscience* 4, no. 2 (2003): 139–47.

Worthy, D. A., K. A. Byrne, and S. Fields. 'Effects of Emotion on Prospection During Decision Making'. *Frontiers in Psychology* 5, no. 591 (2014).

Wright, Robert. 'The Uses of Anthropomorphism' In de Waal, *Primates & Philosophers*, 83-97.

———. 'Why Can't We All Just Get Along? The Uncertain Biological Basis of Morality'. *Atlantic*, 23 October 2013. http://www.theatlantic.com/magazine/archive/2013/11/why-we-fightand-can-we-stop/309525/.

Wynne-Edwards, V. C. *Animal Dispersion in Relation to Social Behaviour*. New York, NY: Hafner, 1962.

Young, L., and J. Dungan. 'Where in the Brain is Morality? Everywhere and Maybe Nowhere'. *Society for Neuroscience 7* (2012): 1–10.

Youssef, F. F., K. Dookeeram, V. Basdeo, E. Francis, and M. Doman. 'Stress Alters Personal Moral Decision Making'. *Psychoneuroendocrinology* 37 (2012):491–98.

Zengotita, T. de. 'Ethics and the Limits of Evolutionary Psychology'. *The Hedgehog Review* 15, no. 1 (2013): 34.

Appendix A

Table A-1: Typology of Environmental Ethics

Normative Grouping	Environmental Ethics	Representative Statement
Anti-Environment	Threat to survival	Nature can be dangerous to human survival.
	Spiritual evil	Nature can be spiritually evil.
Benign Indifference	Storehouse of raw materials	Nature is a storehouse of raw materials that should be used by humans as needed.
	Religious dualism	Humans were created as more important than the rest of nature.
	Intellectual dualism	Because humans can think, they are more important than the rest of nature.
Utilitarian Conservation	Old humanitarianism	Cruelty towards animals makes people less human.
	Efficiency	The supply of goods and services provided by nature is limited.
	Quality of life	Nature adds to the quality of our lives (e.g. outdoor recreation, natural beauty).
	Ecological survival	Human survival depends on nature and natural processes.
Stewardship	Religious/spiritual duty	It is our religious responsibility to take care of nature.
	Future generations	Nature will be important to future generations.
	God's creation	Nature is God's creation.
	Mysticism	All living things are sacred.
Radical Environmentalism	Humanitarianism	Animals should be free from needless pain and suffering.
	Organicism/animism	All living things are interconnected.
	Pantheism	All living things have a spirit.
	Natural rights	All living things have a moral right to exist.

Source: Ben Minteer