Sexual Selection and the Evolution of Morality

A thesis submitted for the degree of Doctor of Philosophy at the Australian National University

by

Benjamin James Fraser

in

March 2010
I hereby certify that this thesis is my own original work.

[Signature]
Short Table of Contents

Acknowledgements viii
Abstract ix

Chapter 1. The Sexual Selection for Morality (SSM) Hypothesis
1.1. Introduction 1
1.2. Presenting SSM 6
1.3. Summary and Preview 39

Chapter 2. The Evolution of Morality: Targets of Explanation
2.1. Introduction 45
2.2. Explanations of Morality 54
2.3. Defending Revised SSM 71
2.4. Conclusion 92

Chapter 3. The Evolution of Morality: Explanatory Mechanisms
3.1. Introduction 95
3.2. The By-Product Explanation for Morality 107
3.3. Natural vs. Sexual Selection 115
3.4. Conclusion 138

Chapter 4. SSM: Conceptual Clarification
4.1. Introduction 141
4.2. Signals and Costs in SSM 143
4.3. Beyond the Handicap Principle 156
4.4. Conclusion 167

Chapter 5. SSM: Empirical Evidence
5.1. Introduction 171
5.2. Clarifying the Commitments of SSM 172
5.3. Assessing the Evidence for SSM 187
5.4. The Many Receivers Problem 202
5.5. Conclusion 206

Chapter 6. SSM: Metaethical Implications
6.1. Introduction 209
6.2. Evolutionary Debunking Arguments 211
6.3. The Reliabilist Reply 222
6.4. Reliability and Moral Cognition 239
6.5. Conclusion 250

Works Cited 253
Expanded Table of Contents

Chapter 1. The Sexual Selection for Morality (SSM) Hypothesis

1.1. Introduction 1
   1.1.1 Evolution and Morality 2

1.2. Presenting SSM 6
   1.2.1. Initial (2000) Version of SSM 7
   1.2.2. Criticisms of SSM Initial (2000) Version 12
      1.2.2a. Explanatory Target 12
      1.2.2b. Greedy Theorizing 15
      1.2.2c. Conceptual Confusions 19
      1.2.2d. Empirical Inadequacy 28
      1.2.2e. Philosophical Implications 34

1.3. Summary and Preview 39

Chapter 2. The Evolution of Morality: Targets of Explanation

2.1. Introduction 45
   2.1.1. Untangling Terminology 46

2.2. Explanations of Morality 54
   2.2.1. Explanations of Behavioural Morality 55
   2.2.2. Explanations of Social Morality 56
   2.2.3. Explanations of Psychological Morality 59
      2.2.3a. SSM and Virtue Ethics 61
      2.2.3b. Revising SSM: Miller plus Joyce 65

2.3. Defending Revised SSM 71
   2.3.1. ‘Joyce-style’ Moral Judgements 72
   2.3.2. The Extent of the Moral Domain 80

2.4. Conclusion 92

Chapter 3. The Evolution of Morality: Explanatory Mechanisms

3.1. Introduction 95
   3.1.1. Evolutionary Explanations for Helpfulness 97
      3.1.1a. Mutualism, Reciprocity, and Kin Selection 97
      3.1.1b. Human Helpfulness and Mistaken Identity 104

3.2. The By-Product Explanation for Morality 107
   3.2.1. Spandrels, Exaptations, and Innateness 110

3.3. Natural vs. Sexual Selection 115
   3.3.1. Darwin’s Distinctions 116
   3.3.2. Are Natural and Sexual Selection Distinct Processes? 122
      3.3.2b. Utilitarian vs. Signal Selection 131
   3.3.3. The Consequences of Collapse 137

3.4. Conclusion 138
### Chapter 4. SSM: Conceptual Clarification

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1. Introduction</td>
<td>141</td>
</tr>
<tr>
<td>4.2. Signals and Costs in SSM</td>
<td>143</td>
</tr>
<tr>
<td>4.2.1. Sorting Signals</td>
<td>143</td>
</tr>
<tr>
<td>4.2.2. Counting Costs</td>
<td>146</td>
</tr>
<tr>
<td>4.3. Beyond the Handicap Principle</td>
<td>156</td>
</tr>
<tr>
<td>4.3.1 Costly Signalling Theories</td>
<td>156</td>
</tr>
<tr>
<td>4.3.2. Signals of Quality vs. Signals of Intent</td>
<td>158</td>
</tr>
<tr>
<td>4.3.3. Revising SSM (again)</td>
<td>160</td>
</tr>
<tr>
<td>4.4. Conclusion</td>
<td>167</td>
</tr>
</tbody>
</table>

### Chapter 5. SSM: Empirical Evidence

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1. Introduction</td>
<td>171</td>
</tr>
<tr>
<td>5.2. Clarifying the Commitments of SSM</td>
<td>172</td>
</tr>
<tr>
<td>5.2.1. Morality and Life History</td>
<td>174</td>
</tr>
<tr>
<td>5.2.2. Morality and Sexual Dimorphism</td>
<td>177</td>
</tr>
<tr>
<td>5.3. Assessing the Evidence for SSM</td>
<td>187</td>
</tr>
<tr>
<td>5.3.1. A case study: Meriam turtle hunting</td>
<td>188</td>
</tr>
<tr>
<td>5.3.2. Competitive Altruism</td>
<td>195</td>
</tr>
<tr>
<td>5.4. The Many Receivers Problem</td>
<td>202</td>
</tr>
<tr>
<td>5.5. Conclusion</td>
<td>206</td>
</tr>
</tbody>
</table>

### Chapter 6. SSM: Metaethical Implications

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1. Introduction</td>
<td>209</td>
</tr>
<tr>
<td>6.2. Evolutionary Debunking Arguments</td>
<td>211</td>
</tr>
<tr>
<td>6.3. The Reliabilist Reply</td>
<td>222</td>
</tr>
<tr>
<td>6.3.1. Conditions for Reliability</td>
<td>229</td>
</tr>
<tr>
<td>6.3.1a. Environment</td>
<td>230</td>
</tr>
<tr>
<td>6.3.1b. Information</td>
<td>232</td>
</tr>
<tr>
<td>6.3.1c. Error</td>
<td>234</td>
</tr>
<tr>
<td>6.3.1d. Tracking</td>
<td>236</td>
</tr>
<tr>
<td>6.4. Reliability and Moral Cognition</td>
<td>239</td>
</tr>
<tr>
<td>6.4.1. SSM and the Tracking Condition</td>
<td>243</td>
</tr>
<tr>
<td>6.5. Conclusion</td>
<td>250</td>
</tr>
</tbody>
</table>
Acknowledgements

Thank you to my supervisors, Kim Sterelny, Richard Joyce, and Brett Calcott, for your expertise and for your patience.

For support and encouragement, thanks to Marc Hauser, Mitchell Joe, Joshua Knobe, Bill Lycan, Geoffrey Sayre-McCord, Ken Perszyk, Luke Russell, and all the graduate students from ANU and UNC-CH who have talked with me about my ideas.

Much love to my dad, Peter Fraser. I could not have done this without you.

To Hatha McDivitt, big big thanks.
Abstract

In this thesis I discuss the Sexual Selection for Morality (SSM) hypothesis, which seeks to explain the evolution of morality in terms of sexual selection and costly signalling. The first part of the thesis puts SSM on the table for critical dissection. My starting point is the work on sexual selection, costly signalling, and morality by evolutionary psychologist Geoffrey Miller. I clarify and elaborate upon Miller’s views by drawing on recent work in signalling theory. The result is a clearer and more nuanced version of SSM. In the second part of the thesis, I evaluate the empirical evidence for this revised version of SSM. I first clarify the predictions of the hypothesis, then survey relevant literature from biology, psychology, anthropology, and economics. Findings from these fields combine to support the claim that moral behaviour plays a signalling role. However, it appears likely that such signals are aimed at many receivers in addition to potential mates, including potential social allies as well as sexual and social rivals. SSM should thus be seen as part of a more general signalling-based account of morality. The third and final part of the thesis considers the metaethical implications of the truth of SSM. A relatively recent arrival on the metaethical scene is the Darwinian debunker, who claims that an evolutionary explanation of morality is, broadly speaking, undermining of morality. A prominent line of reply has been to claim that the fact (if it be one) that human morality is an adaptation shaped by selection over many millennia provides reason to think that our faculty for making moral judgements is likely to mostly produce true judgements. I call this the reliabilist reply. If SSM is true, I argue, then not all the conditions required for the success of the reliabilist reply are met. In particular, what I call the tracking condition fails to be met. The truth of SSM thus counts against the plausibility of the reliabilist reply and lends support to Darwinian debunkings of morality.
Chapter 1. The *Sexual Selection for Morality* (SSM) Hypothesis

1.1. Introduction

When I was a very new graduate student, interested in ethics and the philosophy of biology, an older philosopher told me a story about his unfortunate experience at the intersection of those two fields. This man once taught philosophy of biology to a class of undergraduates. He was explicit at the start of the course about his lack of interest in any ethical implications of the material he planned to present. His was not a course in bioethics. Ethics was not on the menu. Nevertheless, after every class, he was confronted by an especially earnest student, a species familiar to all academics. This particular specimen would invariably ask about the ethical implications of the material just presented. After weeks of increasingly curt dismissals the student was unperturbed. The philosopher was anything but. Finally, after a lecture and before the student could say more than “I was wondering about—”, the philosopher seized the student by the shirtfront and grated “Look mate, *ethics* is for *girls*!” In what follows, I will explore the possibility that this somewhat rash pronouncement may in fact be correct, though not in the sense intended.

Ethics, according to evolutionary psychologist Geoffrey Miller ([2000]; [2007]), *is* for girls. It is also for guys. Miller claims that “morality is a system of sexually selected handicaps,” costly signals that honestly advertise attractive qualities to potential mates (2000: 294). The two key components of Miller’s view about the evolution of morality are thus *sexual selection* and *costly signalling*. Sexual selection was proposed by Charles Darwin ([1859]; [1871]) to explain extravagant ornaments and armaments, such as the peacock’s flamboyant tail and the massive antlers of elk, that serve no function but to increase their bearers’ access to mates. Costly signalling was proposed by Amotz Zahavi ([1975]; [1997]), whose “handicap principle” provided a way of solving the problem of reliability in signalling. I will refer to Miller’s view about the evolution of morality as the Sexual Selection for Morality hypothesis, henceforth SSM. My goal here is to present, clarify, and
evaluate SSM, and eventually to draw out some of its philosophical implications (especially the ethical ones).

1.1.1. Evolution and Morality

The nature of the relationship between evolution and morality has been an issue since Darwin’s presentation of his theory in *The Origin of Species* and *The Descent of Man* right down to the present day. Darwin claimed in the *Origin* that “natural selection acts solely by and for the good of each being” (1998/1859: 648). When it came to applying his theory to our own species in the *Descent*, he believed that “of all the differences between man and the lower animals, the moral sense or conscience is the most important” (2004/1871: 120). Of our moral sense, he wrote:

It is the most noble of all the attributes of man, leading him without a moment’s hesitation to risk his life for that of a fellow-creature; or after deep deliberation, impelled simply by the deep feeling of right or duty, to sacrifice it in some great cause (2004/1871: 120).

To account for the evolution of the human moral sense, Darwin appealed to group selection:

[A] high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe [but] an increase in the number of well-endowed men and an advancement in the standard of morality will certainly give an immense advantage to one tribe over another... At all times throughout the world, tribes have supplanted other tribes; and as morality is one important element in their success, the standard of morality and the number of well-endowed men will thus everywhere tend to rise and increase (2004/1871: 157-158).

Herbert Spencer (1969/1851) agreed with Darwin in thinking that an evolutionary explanation could be given for the human moral sense. Spencer thought in addition that an understanding of evolution provided us with moral guidance. He saw evolution as a progressive and perfecting process:

Their carnivorous enemies not only remove from herbivorous herds individuals past their prime, but also weed out the sickly, the malformed, and the least fleet or powerful. By the aid of which purifying processes, as well as by the fighting, so universal in the pairing season, all vitiation of the race through the multiplication of its inferior samples is prevented; and the
maintenance of a constitution completely adapted to surrounding conditions, and therefore most productive of happiness, is insured (1969/1851: 353).

Spencer thought on this basis that we ought not interfere with the process of evolution.¹ In particular, we ought not adopt welfarist strategies for supporting the less-well-off in society. In Spencer’s view:

The poverty of the incapable, the distresses that come upon the imprudent, the starvation of the idle, and those shouldering aside of the weak by the strong, which leave so many “in shallows and in miseries,” are the decrees of a large, far-seeing benevolence (1969/1851: 353-354).

If Spencer’s position is that evolution provides moral guidance, then Thomas Henry Huxley is his opposite number.² Huxley discussed the relationship between evolution and morality in his Romanes Lectures of 1893. He claimed that “the ethical progress of society depends, not on imitating the cosmic process, still less in running away from it, but in combating it” (2007/1893: 69). A similar sentiment is expressed in Richard Dawkins’ more recent exhortation to “rebel against the tyranny of the selfish replicators [i.e. genes]” and try “deliberately cultivating and nurturing pure, disinterested altruism” (1989/1976: 201).

While Huxley worried that evolution and morality were somehow at odds, Samuel Wilberforce, then Bishop of Oxford, went further, claiming evolution and morality were incompatible. In his 1860 review of the Origin, Wilberforce emphasised the fact that Darwin included humans within the purview of the theory of evolution. Wilberforce used this fact to raise doubts about Darwin’s commitment to Christianity:³

[S]uch a notion is absolutely incompatible not only with single expressions in the word of God on that subject of natural science with which it is not immediately concerned, but, which in our judgement is of far more importance, with the whole representation of that moral and spiritual

¹ At least as social policy-makers. Spencer approved of private charity.
² The phrase that famously if misleadingly encapsulates Darwin’s theory of evolution by natural selection (i.e. ‘the survival of the fittest’) was coined by Spencer, so it seems strangely fitting that the label under which Spencer’s views are vilified (i.e. ‘Social Darwinism’) is not his own but that applied by Marxist historian Richard Hofstadter (1944) to American capitalist industrialists whose allegiance to Spencer’s views was tenuous at best.
³ Confirmed, as it turns out, by what came to light in Darwin’s posthumously published notebooks (see Barrett et al. [2009]).
condition of man which is its proper subject-matter. Man's derived supremacy over the earth; man's power of articulate speech; man's gift of reason; man's free-will and responsibility; man's fall and man's redemption; the incarnation of the Eternal Son; the indwelling of the Eternal Spirit,—all are equally and utterly irreconcilable with the degrading notion of the brute origin of him who was created in the image of God (1860: 258).

Wilberforce can be read as claiming that Darwin's theory, if true, undermines human dignity, of which our capacity for morality is a part. Wilberforce took the incompatibility of evolution and morality to be evident and did not offer arguments to that effect in his review. Others were, if not more rigorous, at least less long-winded in voicing similar concerns. The wife of a different Bishop, upon hearing of Darwin's theory, said "let us hope it is not true or, if it is, let us hope it does not become widely known."

The foregoing gives a sense of the perceived importance of understanding how evolutionary theory impacts on our ideas about morality. It also gives a sense of how muddled much discussion of this issue has been. Whether morality is susceptible to evolutionary explanation, on the one hand, and whether giving such an explanation implies anything about the content or status of morality, on the other, are very different questions. It is vital to keep them distinct. Phillip Kitcher (1985: 417-418) distinguishes four ways in which evolutionary theory might be relevant to morality. Evolutionary theory, Kitcher thinks, might reveal:

- (1) how people come to acquire ethical concepts, to make ethical judgements about themselves and others, and to formulate systems of ethical principles
- (2) facts about human beings that, in conjunction with moral principles that we already accept, can be used to derive normative principles that we had not yet appreciated
- (3) what ethics is all about and [so] settle traditional questions about the objectivity of ethics
- (4) new normative principles

---

4 Dowe (2005: 132) and Pope (2008: 190) provide useful historical context here.
5 The lady was the wife of the Bishop of Worcester. Quoted in Levy (2004: 4).
Darwin was clearly focused on (1). Spencer might be read as pursuing (2) or (4).\(^6\) Wilberforce's worry is probably best located in the vicinity of (3). I am interested here in (1) and (3).

When investigating Kitcher's first possibility, it has become standard to look for insight to work on cooperation and apparent altruism (e.g. Hamilton [1964]; Trivers [1971]; Alexander [1987]; Axelrod [1984]; Sober & Wilson [1998]; Boyd & Richerson [1992]). Such work focuses on showing how particular strategies for interacting with others may be, often despite first impressions, fitness-enhancing either for the individuals employing those strategies or for the groups of which those individuals are members. Much of the interest in Miller's work on SSM lies in the way he shifts focus from the question of how to interact with a given partner to that of with whom to interact and how to be chosen as a partner by others.\(^7\) The question for Miller is less how to interact than with whom to interact. His work is thus representative of a relatively novel way of pursuing Kitcher's first possibility.

Kitcher's third possibility – that considerations drawn from evolutionary theory may reveal something about the status of morality – has recently been investigated by 'Darwinian debunkers.'\(^8\) Such folk claim that successfully realising Kitcher's first possibility, that is, providing an evolutionary explanation for our tendency to make moral judgements, would

---

\(^6\) Robert Richards (1989: 243-74) offers a detailed and more-than-usually sympathetic discussion of Spencer. Richards suggests it was Spencer's prior commitment to a particular utopian social vision that combined with his view of evolution to generate his specific moral proposals (Michael Ruse [2004: 30] makes congruent observations). This is very different to Spencer's view as depicted by G.E. Moore: "the course of 'evolution,' while it shews us the direction in which we are developing, thereby and for that reason shews us the direction in which we ought to develop" (2008/1903: 97). Which reading of Spencer is correct is not as important, I think, as the realisation that even after availing ourselves of Kitcher's framework, identifying which possibility a certain thinker is pursuing may remain a challenge.

\(^7\) As Randolph Nesse puts it, Miller:

...shifts attention away from individual strategies in iterated exchanges, and toward the prior and larger fitness challenges of identifying the best available partners and doing whatever will get them to choose one as a partner (2007: 144).

Miller is one of several researchers pursuing this kind of strategy for understanding the evolution of human cooperation, altruism, and morality. Others include Nesse himself, Ronald Noe (2006), and several more whose work will be discussed in the course of this thesis.

\(^8\) Prominently including Richard Joyce ([2001]; [2006]), Sharon Street (2006), Joshua Greene (2003) and Michael Ruse (1986).
in some way undermine morality. In light of such an explanation, the Darwinian debunkers say, we should see morality as "thoroughly saturated with [distorting] evolutionary influence" (Street [2006: 114]), a "collective illusion foisted upon us by our genes" (Ruse [1986: 27]) that is, at best, a useful "myth" (Joyce [2001: xii]).

One line of reply to the Darwinian debunkers' claims has been to suggest that if human morality is an adaptation shaped by selection over many millennia, then our evolved faculty for making moral judgements is likely to be reliable, to mostly produce true judgements. For example, R. Jay Wallace (2004) writes:

The evolutionary account [of morality] on offer operates through the general human capacity for belief formation, since the account attributes to agents a tendency to form the belief that cooperative behaviours are categorically required. But surely the general human capacity for belief formation is one that does not operate independently of the evidence for or against the beliefs that are held. If evolution is to exploit this mechanism by inducing a disposition to form moral beliefs with a certain content, it is plausible to suppose that it will be able to do this only if the disposition induced does not operate in a way that is independent of the truth. In other words, if we just consider the evolutionary evidence by itself, there seems no good reason to hypothesise that the disposition to form beliefs about categorical moral requirements operates in a way that is independent of the evidence – perhaps no such disposition would have emerged from evolution if the moral belief had been false.

If SSM is correct, I will argue, it provides grounds on which to argue against this kind of reliabilist reply (as I will call it) to Darwinian debunkers. In brief, if SSM is correct, then signalling qualities of oneself rather than tracking features of one's environment was what mattered in the evolution of morality. Or so I will argue when I discuss Kitcher's third possibility in detail.

1.2. Presenting SSM

The remainder of this chapter will be devoted to putting SSM on the table for critical dissection in later chapters. A disclaimer regarding the following material is necessary: Miller's work on SSM is interesting and worth pursuing but in need of much clarification and development. In this chapter I will focus on Miller's initial presentation of SSM in his book *The Mating Mind* (2000),
which was an ambitious attempt to use sexual selection to explain the evolution of art, language, humour, creativity, sportsmanship, morality, and more besides. Miller later (2007) focuses on developing the claim that sexual selection was important in the evolution of morality. I will consider his later work on SSM in due course. Miller’s updated version of SSM is best considered in the light of criticisms of the initial version, though.

In what follows, I will present and offer several criticisms of Miller’s initial case for SSM. Miller’s later development of the view partially addresses some of these criticisms. There is nevertheless still much to do in the way of making SSM conceptually clear and empirically well-supported.

1.2.1. SSM: Initial (2000) Version

Miller begins his initial presentation of SSM with the observation that “any good evolutionary theory of human morality must convert the apparent costs of helping others into a realistic benefit to one’s genes, by turning material costs into survival or reproductive benefits” (2000: 295). He recognises that some of this conversion work can be done by appealing to kin selection (Hamilton [1964]) and reciprocity (Trivers [1971]; Alexander [1987]; Axelrod [1984]) but claims such processes “hardly touch some of the moral virtues we consider most important” (2000: 304). He thinks sexual selection and costly signalling can explain features of human morality kin selection and reciprocity cannot. Miller’s first concern is thus to make plausible the idea that the costs involved in benefiting others can be recouped via access to more and/or better mates. He tries to do this by presenting research from anthropology on big game hunting and from biology on sentinel behaviour in birds.

Kristen Hawkes (1991) claims that big game hunting has several features best explained by treating it as a form of display rather than as

---

9 The unspoken equation here of morality with costly, other-benefiting behaviour has not gone unnoticed. Precisely what the target of an evolutionary explanation of morality should be will be discussed when I critically evaluate SSM.
primarily a provisioning strategy. Her key claims are as follows. Big game hunting is less efficient than hunting smaller game. Controlling the distribution of the spoils of a successful hunt is hard if not impossible, making it difficult for hunters to preferentially direct shares to their family. Receiving meat from a successful hunt is not typically contingent on providing meat oneself in future, making explanations of big game hunting in terms of reciprocity seem implausible. Given these features of hunting, Hawkes suggests that “family provisioning might not always be the only, or even the principal goal of foraging men” (1991: 30). Miller agrees, claiming that big game hunting seen as a means of provisioning appears to be “a risky, wasteful act of altruism” (2000: 311). Seen as a form of showing off, though, big game hunting may make evolutionary sense. Hawkes (1991: 32) suggests that successful hunters benefit both by impressing other males (i.e. potential social allies) and by impressing females (i.e. potential mates). Miller again agrees, claiming that hunting reliably advertises attractive physical and mental qualities, such as strength, endurance and perception, and thus allows the hunter to attract more and/or better mates.

To further bolster the idea that costly, other-benefiting behaviour can benefit the agent via increased mating opportunities, Miller appeals to Amotz and Avishag Zahavi’s (1997) observations of Arabian babblers, a socially-nesting bird species. The Zahavis report that these birds compete to perform risky and/or costly acts that benefit other members of the group, such as standing sentinel at the exposed top of the nest tree or feeding unrelated members of the group. This puzzling phenomenon is explained, they claim, by the fact that birds that perform the most guard duty and feeding also enjoy the highest status within the group and the highest reproductive success (Zahavi & Zahavi [1997: 133-146]).

Miller writes:

I have focused on the apparently altruistic aspects of hunting, not because I am interested in hunting per se, but because it raises a more general issue:

---

10 The show-off hypothesis is controversial. I will take up the controversy in due course.
how could selfish genes possibly give rise to costly, seemingly altruistic forms of charity? (2000: 312).

It thus appears that he takes work on big game hunting and babbler social dynamics to provide a kind of possibility proof for SSM, a demonstration that appealing to signalling and sexual selection to account for puzzlingly costly and other-benefiting behaviours is, at least sometimes, a viable explanatory strategy.

Miller next attempts to establish that, in humans, there is a mate preference for a trait that moral behaviour might plausibly be taken to signal. Miller leans heavily here on the work of evolutionary psychologist David Buss (1989). Buss surveyed the mate preferences of more than ten thousand people, men and women, from 37 countries. He found that the most desired trait in a potential long-term mate, across cultures and sexes, was kindness (1989: 13). Miller uses Buss’ work to support SSM as follows. If there is a culturally universal preference for kindness in potential long term mates, then it is plausible that displays of moral virtue evolved via inter-sexual selection, since such displays can serve as reliable signals of kindness, and that quality is very important in mate choice. In sum, by establishing that we prefer kind mates, Miller aims to make it plausible that there has been sexual selection for some kind of display that advertises kindness; he thinks much moral behaviour fits that bill.

It is worth taking a moment to consider a criticism of Miller from Catherine Driscoll (2006: 515), who says “it is not clear that the ‘kindness’ referred to in Buss’ study is “altruism” in the biological sense…nor is it

---

11 Much of the controversy surrounding Buss’ work (and that of evolutionary psychologists conducting similar research) has focussed on claims about sex differences in mate preferences. Claims that men more so than women value youth and beauty in a mate, while women have a stronger preference than men for wealthy, high-status partners, have attracted plenty of critical attention (e.g. Buller [2005]; see Machery & Barrett [2006] for a reply). Buss’ finding that both sexes agree on the paramount importance of kindness has received comparatively little attention, being perhaps not incendiary enough to generate much interest. As far as SSM is concerned, though, that is Buss’ most important finding. That finding has since been confirmed by other researchers (e.g. Li et al. [2002]). I will therefore largely ignore the fireworks sparked by Buss’ claims about sex differences in mate preferences.

12 I am reconstructing here. Miller (2001) has said his discussion of morality in The Mating Mind is meant to “take seriously” Buss’ findings but he does not explicitly lay out his reasoning on how these findings support SSM.
obviously the extreme, showy altruism Miller wants to explain.” Driscoll thinks it is therefore not clear what support Buss’ study lends to Miller’s view. Driscoll is right that showy altruism and kindness are not the same thing, but that is not Miller’s claim. Miller’s claim is that showy altruism, including conspicuous moral behaviour, is a reliable signal of kindness. Driscoll’s criticism is based on a confusion between the (supposed) signal and the (supposed) quality signalled.

When Miller directly discusses morality, his approach is to do some creative story-telling about how a variety of phenomena, which he takes to be elements of morality, might serve as signals to potential mates. I will focus here on his treatments of charity and courtship gifts, sympathy, and fairness and sportsmanship.13

Miller claims that modern charitable giving has features that fit poorly with the idea that the point of the exercise is to effectively help those in need. The first is the inefficiency of many charities, as measured by the percentage of donated funds that goes to helping the needy as opposed to running and promoting the charity itself. The second is the premium placed on publicising donations and causes, evidenced by the badges and ribbons given to donors and by the reluctance of donors to support obscure but otherwise worthy organisations.14 Miller suggests that “the point of charity is to incur the cost of giving rather than to bring real benefits to others,” that donating to charity is “a form of showy, wasteful display” (2000: 324).15

Courtship gifts are often expensive but useless, flowers and jewellery being two cases in point. Miller asks how evolution could have “favoured humans who fall in love with individuals who provide them with useless luxuries that bring no survival benefits” (2000: 329). Miller’s answer, unsurprisingly, proceeds in terms of sexual selection and costly signalling.

---

13 Miller also talks about tipping, romantic love, paternal solicitude and visionary leadership. I am focusing here on the most substantial elements of his discussion.
14 How much these observations support treating charity as signalling will be discussed later in this chapter.
15 This is not to say Miller thinks those who devote money or time to charity do not care about the needy or are aiming to show off.
The utility of romantic gifts is not as important as the cost they incur. Miller suggests that gift-givers signal their willingness and ability to invest in a potential relationship by incurring the cost of a courtship gift.

Sympathy, according to Miller, is “the emotional capacity to care about [others’] suffering” (2000: 330). He observes that “much of human courtship consists of sympathy displays” and that “the development of emotional intimacy could be viewed as the mutual display of capacities for extremely high levels of sympathy” (2000: 330). Miller tells two stories about sympathy, signalling, and mate choice. On the first and simpler story, sympathetic displays (whatever form they take) are signals of sympathy, which is a desirable quality in potential partners. The second story starts with the claim that, as mate choices go, choosing a psychopath is a poor one. On this story, “sympathy indicators evolved to advertise freedom from psychopathy,” since “[psychopaths] lack the sympathy and agreeableness that average people have” (2000: 331).

Miller’s favoured explanations for our sense of fairness and our inclination to participate in and/or pay attention to competitive sports are intertwined. In Miller’s view, sport is a ritualisation of sexual and social competition. Playing sports allow participants to display their abilities and spectators to compare players. Ideally, winners in sports are determined meritocratically. Miller notes that “fair competition maximises the information that winning carries about the relative fitness of the winner” (2000: 336). Miller assumes that competitive sporting performances play at least some role in mate choice and, although we value fairness in many arenas other than sport, he claims that “it is likely that our concerns for fairness evolved in this context [i.e. mate choice]” (2000: 337).16

---

16 One might well wonder whether competitive sport is a suitable target for evolutionary explanation. Is it a human universal? How deep does its history go? Miller includes club-fighting amongst Yanomamo tribesmen as a sport, indicating that he has a rather liberal definition of the term on which playing sport is not a culturally idiosyncratic and relatively modern pastime. Of course, whether Miller’s preferred evolutionary explanation for the practice is correct is another matter.
Miller concludes his initial presentation of SSM by dismissing the worry that the truth of SSM would somehow cheapen or debase morality. He repeatedly emphasises that “a sexual function is not a sexual motivation” (2000: 320, 321, 325, 340). Miller is quite right to dismiss this worry. Just as talk of selfish genes does not imply that we never care about others for their own sake, so talk of sexually-selected morality does not imply lasciviousness on the part of moral agents. With this point made, Miller is content to “leave it to others to discuss the existential and theological implications of living in a lineage that directed its own evolution through its powers of mate choice” (2001).

1.2.2. Criticisms of SSM Initial (2000) Version

As is likely evident, Miller’s initial presentation of SSM requires much in the way of conceptual clarification and empirical bolstering. I will now offer several criticisms of Miller’s initial presentation of SSM. My aim in doing so is to set up the issues to be discussed in later chapters of this thesis. My criticisms will concern Miller’s target of explanation, his view of the relationship between sexual selection and other processes in the evolution of morality, his neglect of several important distinctions in signalling theory, the quality and quantity of his evidence for SSM and, finally, his discussion of the broader implications of the truth of SSM.

1.2.2a. Explanatory Target

Miller certainly takes himself to be explaining human morality. He mentions moral behaviour, moral judgement, moral emotions, moral reasoning, moral instinct, moral motivation, and moral virtue, as well as charity, sympathy, generosity, and fairness. He is not particularly careful in defining ‘morality’ as a target of explanation, though. The definition he does offer – “any behaviour that displays good moral character” (2000: 293) – is unhelpful. Furthermore, while being charitable and fair are at least paradigm cases of moral behaviour, romantic gift giving and excelling at sport are not, meaning Miller’s conception of morality is not even particularly intuitive. Rather than seeing
Miller as explaining one kind of behaviour – moral behaviour – one might be tempted to see him as offering a set of structurally similar explanations (all appealing to sexual selection and costly signalling) for a variety of different human behaviours, only some of which count as moral.

Miller anticipates this line of criticism. He claims the problem lies with the critics. Most moral philosophers and evolutionary psychologists narrow-mindedly “consider only a tiny fraction of human virtues and moral judgements worthy of analysis” (2000: 339). Miller recommends broadening the scope of evolutionary investigations of morality to include Nietzsche’s pagan virtues, among which Miller counts bravery, skill, beauty, fertility, strength, pride, and leadership.

It is true that different philosophers and cultures have held quite disparate views about what kinds of behaviour are morally required, praiseworthy, and so on. Miller’s reply, though, misses the deeper point of the criticism. The problem is not whether the behaviours Miller is explaining by appeal to sexual selection and costly signalling should be counted as moral. The problem is that focusing on kinds of *behaviour* in giving an evolutionary explanation of morality is a mistake. Miller overlooks a crucial distinction between two things that might be meant by the claim that human morality is the product of evolution:

...that we naturally act in ways that are morally laudable – that the process of evolution has designed us to be social, friendly, benevolent, fair, and so on... [or] that the process of evolution has designed us to think in moral terms (Joyce [2006: 3]).

The first claim is a claim about behaviour while the second is a claim about a mechanism for producing behaviour. Miller’s initial presentation of SSM is clearly aimed at showing how the claim that human morality is the product of evolution could be true in the first sense. Succeeding in this endeavour would not establish that the claim is true in the second sense, however. To see why, consider for a moment a non-human species, such as the
naked mole rat (*Heterocephalus glaber*).\(^{17}\) Suppose we had a complete account of the fitness benefits that explain why naked mole rats have evolved to feed, defend and work for the good of the other members of their group. Feeding, defending, and working for the good of others are all behaviours that we think are morally laudable. Explaining why naked mole rats evolved to perform such behaviours, however, does not establish that *they* think anything of the sort. Of course, naked mole rats do not think anything of the sort, and this is just the point. It would be obviously misguided to explain why naked mole rats perform the kinds of behaviours just mentioned and call that an explanation of the evolution of mole rat morality in the second sense above. Likewise, even if sexual selection and costly signalling can reveal “the hidden evolutionary benefits” (Miller [2000: 306]) of behaviours we think are morally laudable – being kind, generous and the like – that would not itself tell us why we think of such behaviours in moral terms.

To see why the foregoing matters in the bigger picture here, recall Kitcher’s framework for relating evolutionary theory to morality. Kitcher suggests that evolutionary theory might reveal how we came to be the kind of creatures that have moral concepts, make moral judgements, and formulate moral systems. When it comes to realising this possibility, as Richard Joyce (2006: 50) puts it:

> [A]n explanation of how natural selection might end up making human beings with altruistic, sympathetic, loving tendencies toward each other – how, that is, it might produce *nice* humans (or, if you prefer, *virtuous* humans) – misses the target.

That is not to say Miller’s efforts in his initial presentation of SSM are wasted. Appealing to sexual selection and costly signalling to reveal the hidden evolutionary benefits of conspicuously kind or generous behaviour is a novel and worthwhile project. Conspicuously kind and generous behaviors are interesting explanatory targets in their own right, but they are not *the* target if the goal is to realise Kitcher’s first possibility.

\(^{17}\) For a natural history of this surpassingly ugly eusocial rodent, see Sherman *et al.* (1991).
1.2.2b. Greedy Theorising

In his initial presentation of SSM, Miller appears guilty of overemphasising the importance of sexual selection and underplaying the contribution that other processes can make to explaining the kinds of behaviour he is interested in.\(^{18}\) Whether this charge sticks is debatable but Miller is, at the very least, ambivalent about the relationship between sexual selection and other processes, such as kin selection and reciprocity, in the evolution of morality. As noted above, he claims that sexual selection and costly signalling can explain aspects of human moral behaviour that kin selection and reciprocity cannot, for example, helpfulness toward unrelated individuals who are highly unlikely to ever reciprocate the help.\(^{19}\) In some places, Miller allows that “sexual choice does not account for all of morality... Kinship and reciprocity, too, were very important” (2000: 339). Elsewhere, though, he claims that “sexual selection may cast new light on certain moral phenomena that were previously understood in terms of kinship and reciprocity” (2000: 304). Overall, Miller does give the worrying impression of being an explanatory pluralist only in the sense derisively described by Gould & Lewontin (1979: 585):

In natural history, all possible things happen sometimes; you generally do not support your favoured phenomenon by declaring rivals impossible in theory. Rather, you acknowledge the rival, but circumscribe its domain of action so narrowly that it cannot have any importance in the affairs of nature. Then, you often congratulate yourself for being such an ecumenical chap.

To be fair, it is unclear whether Miller really is guilty of greedy theorising. That very lack of clarity is itself a problem, though. Just how sexual selection relates to other processes that are thought to have been important in the evolution of morality needs more careful and detailed

---

\(^{18}\) One reviewer of *The Mating Mind* charged Miller with “greedy theorizing” (Wagner [2002]) while another said that “in reviving sexual selection from its status as second fiddle to natural selection, [Miller] drives the remainder of the orchestra cowering into the aisles” (Horvath [2000]).

\(^{19}\) Miller is unimpressed by explanations of such behaviours as the effect of instincts for kinship and reciprocity operating in evolutionarily novel settings where social networks extend far further than was typical throughout most of human history. I will take up this discussion in Chapter 3 (see §3.1.1b).
consideration than Miller gives in his initial presentation of SSM. Furthermore, Miller devotes the bulk of his discussion to kin selection and direct reciprocity, but these are not the only other processes that must be considered. Indirect reciprocity, mutualism, cultural group selection, and the idea that morality may be in some way an evolutionary by-product all need to be considered in judging the relative merits of SSM as a hypothesis about the evolution of morality.

Relatedly, Miller’s view of sexual selection as a special kind of process distinct from natural selection is problematic. Miller makes much of the supposed distinctness of sexual from natural selection, promoting SSM as a novel alternative to “survival-based” accounts of the evolution of morality (2000: 66). Whether there is a plausible, principled distinction to be drawn between sexual selection and natural selection is, however, highly debatable. Miller adopts Darwin’s view on this matter, in defiance of more recent treatments of evolutionary theory that subsume sexual under natural selection.20

According to Darwin, the natural world poses two explanatory puzzles. One is how to account for the way in which organisms are often impressively well-suited to their environments and lifestyles. The other is how to account for the fact that organisms sometimes bear traits that seem quite ill-suited to their environments and lifestyles, such as extravagant, conspicuous ornaments and elaborate, cumbersome armaments. Further complicating the second puzzle, ornaments and armaments are often present in one sex of a species but not the other. Darwin’s solution to the first puzzle was natural selection. His solution to the second was sexual selection. He distinguished the two processes on the grounds that:

...[sexual selection] depends, not on a struggle for existence in relation to other organic beings or to external conditions [as does natural selection], but

---

20 In her discussion of the history of sexual selection theory, Helena Cronin (1991: 236) writes that the debate over the distinctness or otherwise of sexual and natural selection is nowadays seen as a “a storm in a teacup”, since from the gene’s-eye view, the only question that matters is how survival and reproduction contribute to the replication of genes. Whether the storm has entirely blown itself out is a question I will consider in Chapter 3 (see §3.3).
on a struggle between the individuals of one sex, generally the males, for the possession of the other sex. The result is not death to the unsuccessful competitor, but few or no offspring (1998/1859: 117).

That Darwin saw natural selection and sexual selection as distinct processes is evidenced by numerous quotes. Another runs:

Sexual selection depends on the success of certain individuals over others of the same sex in relation to the propagation of the species; whilst natural selection depends on the success of both sexes, at all ages, in relation to the general conditions of life (2004/1871: 398).

Miller’s own attempts at distinguishing natural from sexual selection are unconvincing. One proposal Miller makes is that sexual selection differs from natural selection in that the former involves active choice on the part of organisms whereas the latter involves only passive filtering of phenotypes by an unintelligent, uncaring environment (2000: 8-10).

This is unconvincing as a way of defending Darwin’s distinction. Selection involving the kind of active choice by organisms that Miller has in mind occurs in contexts other than competition for mates. For example, predators choosing prey or primates engaged in coalition formation “apply their faculties of perception, cognition, memory, and judgement to pick”

21 Miller is not alone in making this kind of proposal for distinguishing natural from sexual selection. Michael Ghiselin (2006) writes:

Ghiselin thinks artificial selection involves conscious human choice, while natural selection involves purely metaphorical selection, which is really just passive filtering of variants by environments, and sexual selection is “conceptually…more similar to artificial selection than to natural selection,” since it involves selecting agents (namely, potential mates) that actually have preferences and make decisions. Ernst Mayr has made the same suggestion:

There is…an important distinction between natural selection and sexual selection. In sexual selection the female is actually selecting males for certain traits, and this is different from natural selection [in which] certain individuals survived because they had certain characteristics, but they weren’t selected. The process consists of eliminating all the others (Mayr as quoted in Shermer & Sulloway [2000]).

…natural selection, taken strictly, is not a selection process at all, but rather a process of elimination. It is the least well adapted individuals that are eliminated in every generation, and those that are better adapted have a greater chance to survive. … Of course, during sexual selection real selection takes place (Mayr [2002]).

Miller’s proposal, then, is ultimately unsuccessful but not entirely idiosyncratic.
(Miller [2000: 9]), not the best sexual partners they can, but the best targets and team-mates, respectively. An organism’s predators may possess at least as much cognitive sophistication as its conspecifics, but even if those predators choose to prey on the weak and sick on purpose (so to speak), to suppose this interaction falls into its own evolutionary category (one other than natural selection) seems silly.

Moreover, this proposal for marking sexual selection off from natural selection would seem to depend on being quite liberal (too liberal, one might think) with the notion of choice. After all, sexual selection supposedly operates even in cognitively unsophisticated organisms. Among stalk-eyed flies, for example, females mate more frequently with wide eyespan males than with narrow eyespan males, a phenomenon that Chapman et al. (2005: 533) consider “a canonical example of sexual selection.” Whatever female flies are doing here, however, is far from a canonical example of active, judicious choice. It is controversial how much cognitive sophistication we can attribute to non-human organisms, but seeing the problems with Miller’s proposal does not require taking any extreme stance on that issue.22

Another proposal Miller makes for distinguishing sexual from natural selection is that the former is more “powerful” than the latter in virtue of operating via positive feedback loops (2000: 69). He writes:

Imagine the headaches if natural selection worked that way [i.e. via positive feedback loops]. Organisms would select which environments exist, as well as environments selecting which organisms exist...Would the feedback loop between polar bears and Arctic tundra result in a tundra of Neptunian frigidity where bears have fur ten feet thick, or a tundra of Brazilian sultriness where bears run nude? (2000: 68).

This too is an unconvincing way of defending Darwin’s distinction. Selection involving positive feedback loops occurs in contexts other than

---

22 Finally, all this is to say nothing of the possibility of sexual selection in plants. Mary Wilson ([1979]; [1990]; Wilson & Burley [1983]) argues that sexual selection operates in plants. Whether or not sexual selection operates in plants is certainly controversial (see Grant [1995] for a review), the very fact that there is controversy here should make us wary of ruling out sexual selection in plants by definitional fiat (as Miller’s proposal would do).
competition for mates, for example, in evolutionary arms races between pathogens and hosts. Furthermore, the kind of scenario Miller paints in such an incredulous light with his polar bear story is actually not so far-fetched. Take the case of the sword-billed hummingbird (Ensifera ensifera) and the banana passion flower (Passiflora mixta). The extremely long, thin bill of the bird and the extremely deep, narrow floral tube of the plant have co-evolved via exactly the kind of positive feedback loop Miller seems to think is distinctive of sexual selection (Lindbergh & Olesen [2001]). Miller’s second suggestion thus also fails to demarcate sexual from natural selection.

Given the problems with Miller’s proposals for distinguishing sexual from natural selection, two questions need to be asked. First, is there some plausible, principled distinction to be drawn between sexual selection and natural selection? Second, to what extent does it matter to SSM whether or not such a distinction can be drawn? I will address these questions in the course of my later discussion (see Chapter 3, especially §3.3.2).

1.2.2c. Conceptual Confusion

Sexual selection is one of the two key components of SSM. Costly signalling theory is the other. Miller relies on Amotz Zahavi’s (1975) statement of the handicap principle in his initial presentation of SSM. Zahavi proposed the handicap principle as a solution to the problem of reliability in biological signalling systems. Before discussing Miller and SSM, I will briefly describe the problem of reliability and Zahavi’s proposed solution to it.

Organisms often find occasion to be interested in unobservable qualities of other organisms. Sexual rivals want to know each others’ fighting ability. Parents want to know how hungry their offspring are. Predators literally on the prowl want to know the evasive ability of potential prey. Organisms metaphorically on the prowl want to know the quality of potential

---

23 It would be insulting to suggest that Miller is unaware of this fact. The problem is that he apparently misses its implications for his proposed distinction between sexual and natural selection.
mates. Making an adaptive decision (about, for example, who to fight, who to flee from, who to feed on and who to mate with) depends crucially on estimating these unobvious qualities. As it turns out, potential partners, predators and prey themselves often provide the relevant information: they roar, beg, stot, sing or dance, for example, or signal in some other way. The problem of reliability arises when we ask why, given the often strong incentive to mislead signal receivers, signal senders do not do so. Why do signallers not present themselves as stronger, faster, hungrier, or more desirable than in fact they are? While there is not always an incentive to deceive in biological signalling systems, such an incentive exists often enough that there is a real puzzle concerning the persistence of signalling systems. Some explanation is required for why signalling systems do not collapse as a result of receivers eventually ignoring a cacophony of dishonest signals.

Zahavi’s solution to the problem of reliability was to suggest that signals could remain trustworthy even in the face of incentives to deceive if signals of a particular quality were too costly for individuals lacking that quality to produce. For example, consider the peacock’s tail.24 Zahavi claimed that an extravagant and cumbersome tail is a reliable signal of high quality because only a high quality peacock can afford to pay the costs of developing and displaying it. In Zahavi’s first formulation of the handicap principle, costs keep signals honest by imposing a test on signallers that only genuinely high-quality individuals can pass.25

Despite a sceptical reception (e.g. Maynard-Smith [1976]), Zahavi’s handicap principle eventually gained acceptance after mathematical modelling work by Alan Grafen ([1990a]; [1990b]) and Rufus Johnstone (1995) showed its feasibility in principle. Miller rightly points out that costly signalling,

24 In correspondence with a cousin Darwin wrote:

I remember well the time when the thought of the eye made me cold all over, but I have got over this stage of the complaint, and now small trifling particulars of structure often make me very uncomfortable. The sight of a feather in a peacock’s tail, whenever I gaze at it, makes me sick!” (Burkhardt et al. [1993: 140])

25 Later work by Zahavi and others refined and redefined the handicap principle. Dealing with these complications and their implications for SSM is work for a later chapter (see §4.3.1).
especially in the context of mate choice, is now one of the most fertile fields in biology. Miller’s adoption and employment of Zahavi’s ideas, however, is not without its problems.

The confusions I will draw attention to are certainly not all Miller’s fault. The field of signalling theory is rife with potential confusions. Some researchers use the same term to mean different things, while others use different terms to mean the same thing. In my discussion of signalling, I will draw on recent treatments of costly signalling theory by John Maynard-Smith & David Harper (2003) and William Searcy & Stephen Nowicki (2005). Both of these works explicitly aim to bring some clarity to the field.

Miller’s initial presentation of SSM pays insufficient attention to several important distinctions in signalling theory, between: signals and cues, expenditures and costs, absolute and relative costs, and handicaps and indices. I will now present these distinctions and explain why they matter when evaluating SSM.

Maynard-Smith & Harper (2003: 3) propose three conditions an act or structure must satisfy to count as a signal. A signal, on their view, is an act or structure that:

(1) alters the behaviour of other organisms,
(2) evolved because of that effect, and
(3) is effective because the receiver’s response has also evolved.

In the course of their discussion, Maynard-Smith & Harper also offer a fourth defining feature of signals. Whether this fourth feature is meant to follow from the three above is unclear. I will treat it as a further, distinct condition that an act or structure must meet in order to count as a signal:

(4) [the act or structure] must be able to evolve independently of any quality of the signaller about which it conveys information (2003: 4).

In what follows, these four conditions will be explained, and several examples detailed, in the course of distinguishing signals from what Maynard-
Smith & Harper term “cues”: “feature[s] of the world, animate or inanimate, that can be used by an animal as a guide to future action” (2003: 15).

Although both signals and cues can feature in the adaptive decision-making and thus affect the behaviour of other organisms, only the former evolved because of that effect. A word of caution is needed here. Maynard-Smith & Harper’s talk of “distinguishing” signals from cues is potentially misleading, inviting as it does the thought that some act or structure could be one or the other but not both. Given the definitions above, all signals are cues but not all cues are signals. Signals are thus a more specific category of acts and structures within the more general category of cues. Some examples will help make clear the difference between signals and cues.

Gillies (1980) found that mosquitoes locate prey by moving up CO₂ gradients in the air. Animals exhaling CO₂ provide mosquitoes with a guide to action when it comes to locating prey. Obviously, though, we would not explain why animals exhale CO₂ by citing the usefulness of their doing so to blood-sucking insects. Exhaling animals provide mosquitoes with a cue but are such animals not signalling to the mosquitoes.

The phenomenon of aposematism (i.e. warning coloration) provides many examples of signals (Grimaldi & Engel [2005: 603]). The patterns on the wings of the poisonous monarch butterfly (Danaus plexippus) affect the behaviour of other organisms, such as the North American bluejay (Cyanocitta cristata), which will not eat another monarch after having been unfortunate enough to sample one. The monarch’s wing patterns evolved because of their effect on predator responses and predator responses also evolved, with selection favouring predators who avoided the poisonous prey. Notice, wing patterns can evolve independently of poisonousness. For example, the African mocker swallowtail (Papilo dardanus) is a non-poisonous species of butterfly that closely resembles other, poisonous butterfly species.²⁷

²⁶ To do so would be a kind of perverse, inverse Panglossianism.
²⁷ This is Batesian mimicry. The classic example of this phenomenon was, until fairly recently, the resemblance between the above-mentioned monarch butterfly and the supposedly
Notice, finally, something can be a signal to one organism and a cue to another. Male tungara frogs (*Physalaemus pustulosus*) give calls consisting of a 'whine' and several 'chucks,' with the number of chucks a male makes increasing with his body size. Since female tungara frogs prefer to mate with larger males, male 'chucking' has been interpreted as a signal of body size meant to attract females (Ryan & Rand [1993]). Unfortunately, larger male frogs are also attractive to fringe-lipped bats (*Trachops cirrhosus*), as prey. These bats preferentially target frogs producing numerous 'chucks.' Thus, 'chucking' can also be seen as a cue to bats.

Two implications of the above distinction between signals and cues must be emphasised before using it to think about Miller's initial presentation of SSM. First, showing an act or structure to be a cue does not provide an explanation for why that act or structure evolved. Signals evolved because of their effect on receivers' behaviour. Cues, despite figuring in other individuals' decision making, evolved for some reason other than their effect on other individuals' behaviour.²⁸

The second implication is as follows. A signal is of interest to an organism facing an adaptive decision because it reveals some further, less immediately apparent feature(s) of the signaller, such as poisonousness in the butterfly example given above. A cue, by contrast, need not be revelatory in this way in order to serve as a useful guide to future action. Cues may themselves be the very features of interest to organisms facing adaptive decisions. The following story illustrates this point.

Suppose a mugger is lurking in the mouth of an alley, watching pedestrians pass close by and wondering who to haul into the shadows. A good thing for the mugger to attend to would be the attire of passers-by. Now

---

²⁸ If they evolved at all. Recall, a cue is any feature of the world that serves an organism as a guide to future action. A cue might be an adaptation (either an act or a structure) borne by an organism, or it might be a brute feature of the abiotic environment that does not have an evolutionary history.

non-poisonous viceroy (*Limenitis archippus*). As it turns out, the viceroy is also poisonous (Ritland & Brower [1991]), so the resemblance is actually a case of Mullerian mimicry.
suppose a person were to walk past the alley swathed in fur and festooned with jewels. Such attire doesn’t signal wealth; it is wealth.

In the story above, a feature of the world serves as a useful guide to future action without indicating some further, distinct feature of the world. This is possible in the case of cues. It is not possible in the case of signals. Signals are always signals of something.

Having distinguished signals from cues, we can use this distinction in thinking about Miller’s initial presentation of SSM. Some of the actions Miller discusses, like conspicuous charity or lavish romantic gift-giving, might reliably indicate a further, less directly observable, feature of the agent. In those cases, the most plausible candidate for that feature is wealth. Other of the actions Miller discusses, like displays of sympathy or fairness, seem to promise only more of the same in future. Sympathy and fairness in a potential partner seem like the very qualities of interest to an individual making a mating decision, rather than indicators of some further, hidden qualities. A signalling story here seems unnecessary. In cases like this, SSM is better stated in terms of cues rather than signals. Some moral behaviours may not reveal any further quality but just themselves be good things to pay attention to when choosing a mate, that is, cues.29

Although it may not appear so, the above would actually constitute quite a drastic revision of Miller’s view. Miller presents sexual selection as an explanation for both why we perform moral behaviours (the supposed signals) and why we prefer such behaviours (our response to the supposed signals). However, moral behaviour not being ignored in mate choice is not the same as it being the product of mate choice. Whether a potential mate has legs or not is certainly not ignored by individuals facing mating decisions, but that does not mean the having of legs is to be explained in terms of inter-sexual selection. Moreover, if paying attention to whether a potential mate is kind or cruel is just a “Good Trick” explicable in terms of the “general nonstupidity” of our

29 There are complexities here that will be examined in detail later (see §4.3.2). The goal for now is simply to raise issues for discussion.
species (Dennett [1995: 485]), then there seems to be no need to posit a special process of sexual selection to explain the widespread preference for kind mates. If SSM is to be plausibly maintained, it is necessary to sort out signals from cues and get clear on the qualities supposedly signalled by various moral behaviours.

I will move on now to the issue of signal costs. A rigorous treatment of the issue of signal costs is lacking in Miller’s initial presentation of SSM. Some of the behaviours Miller discusses clearly demand resources. For example, conspicuous displays of charity and romantic generosity straightforwardly require one to relinquish material wealth. Other of the behaviours he discusses are not so obviously demanding. Sympathy is a case in point. Showing sympathy toward a potential mate (or toward some third party while a potential mate observes) seems to involve few costs aside from the time taken to do so (time that might have been spent doing something else instead). If Miller is to maintain his costly signalling story about the evolution of sympathy, he must make the case that being sympathetic is something that only genuinely sympathetic folk can afford to do. Miller does not make such a case in his initial presentation of SSM and, on the face of it, the prospects for doing so seem poor.

One might think Miller’s case is intuitively sound even if he does not give a precise accounting of the costs of moral behaviour, since acting morally often requires some kind of sacrifice on the part of the agent. There is, however, a distinction between expenditures and costs (Kotiaho [2001: 366]). Expenditure involves using up some resource (be it energy, time or material of some kind) or incurring some kind of risk (for example, predation risk). Cost, at least in the context of costly signalling theory, involves more than expenditure. In Grafen’s (1990) mathematical model demonstrating the in-principle possibility of costly signalling, signalling was costly in the sense that it reduced one component of the signaller’s fitness (for ease of imagination,

30 One might think there is a reason why only genuinely sympathetic folk can put on sympathetic displays but that the reason has not to do with costs but with the difficulty of convincingly faking such displays. This is a point I turn to shortly when discussing the difference between handicaps and indices.
suppose the signal to be a peacock’s tail and the cost to be decreased viability due to increased predation risk). Expenditures may inflict costs, or they may not. That moral behaviour requires expenditures is obvious; all behaviours require expenditures. That those expenditures inflict fitness costs on signallers is a further claim that must be made explicit and defended. Miller’s reliance on a loose, vernacular notion of cost in his initial presentation of SSM is thus problematic.

There is one further issue regarding to signal costs to be raised. Miller’s initial presentation of SSM pays insufficient attention to the fact that signals must be not only costly but differentially costly to ensure reliability in models of costly signalling (again, see Grafen [1990a]; [1990b]; Johnstone [1995]). It is not the absolute cost of a signal that matters in such models but the relative cost of that signal for honest versus dishonest signallers. For cost to ensure reliability, a signal must be more costly for individuals lacking the relevant quality than for those possessing it and so costly that it is not worth the former’s while to dishonestly signal anyway. The tempting benefits of signalling dishonestly must be outweighed by the cost of giving the deceptive signal.

Both Maynard-Smith & Harper (2003: 19-20) and Searcy & Nowicki (2005: 11-13) emphasise the point above. The point seems to go under-appreciated in Miller’s initial presentation of SSM. It cannot simply be assumed that the kinds of behaviour Miller discusses – being charitable, generous, fair, and the like – are differentially costly. Again, Miller’s reliance on a loose, vernacular notion of cost creates problems when SSM comes in for critical scrutiny. SSM needs to be situated more carefully in relation to the treatments of costly signalling in the biological literature.

The final distinction I will raise here is between two kinds of signal, a handicap and an index. Handicaps in Maynard-Smith & Harper’s terminology are costly signals as described above, signals that are kept reliable by the fact that faking is unprofitable. Miller frames SSM in terms of handicaps: “morality is a system of sexually selected handicaps – costly indicators that
advertise our moral character” (2000: 294). Although suitably differential signal cost is one way to maintain reliability in signalling systems, it is not the only way. An index in Maynard-Smith & Harper’s terminology is “a signal whose intensity is causally related to the quality being signalled, and which cannot be faked” (2003: 15). Indices are signals that are reliable because constraints prevent would-be fakers from signalling dishonestly.

Maynard-Smith & Harper use Riechert’s (1978) study of territorial disputes in funnel-web spiders (*Agelenopsis aperta*) to illustrate the distinction between handicaps and indices. Riechert found that spiders confronted with a rival more than 10% bigger than them would withdraw without a fight. Spiders signal their size to each other by vibrating the web on which the confrontation takes place. Vibrating the web requires a miniscule expenditure of energy. The signal is nevertheless reliable because the extent to which a spider is able to vibrate the web is dictated by the spider’s size. For a petite spider, faking heftiness is not an option.31

Having distinguished handicaps from indices, we can use this distinction in thinking about Miller’s initial presentation of SSM. As alluded to earlier, it may be that some of the things Miller discusses are better understood as indices than as handicaps. For example, consider displays of sympathy. Imagine listening to one’s friend relate the woes of her day and, in response, pulling a pained face while touching her hand and murmuring ‘c’est la vie.’ The expenditure demanded by such a performance is minimal. We are nevertheless quite good at telling when displays like this are genuine and when they are not. The minitiae of expression, posture and tone convey reliable information about others’ psychological states. The figure in signalling theory to whom to look for insight here is not Zahavi but economist Robert Frank (1988). Frank emphasised the difficulty of convincingly faking emotions such as anger and, more to the current point, compassion.

---

31 Whether Maynard-Smith & Harper’s distinction between handicaps and indices holds up is debatable. The crucial question is whether there is a real distinction to be drawn between costs and constraints, or whether the latter can be recast as a type of cost (perhaps a developmental cost). This issue will be discussed in detail later (see §4.2.1).
Some elements of what Miller takes to be human morality are best understood as indices rather than as handicaps. This may appear a niggling revision of Miller’s idea but it is quite significant. The distinction between handicaps and indices matters when it comes to providing empirical support for SSM. Too narrow a focus on Zahavi’s handicap principle would result in efforts wasted looking for the differential fitness costs of supposed moral signals when what should instead be looked for are constraints that make faking such signals difficult if not impossible.

In the foregoing discussion, I have drawn several distinctions that go largely unrecognised in Miller’s initial presentation of SSM. I have also indicated how drawing these distinctions matters when evaluating SSM. It is perhaps unfair to try to get such critical mileage out of Miller’s discussion of sexual selection, costly signalling, and morality in *The Mating Mind*. That book was in large part programmatic, describing and advocating a novel approach to understanding the evolution of morality rather than developing the approach in detail. Nevertheless, before that latter task can be attempted, it is important to construct the requisite theoretical machinery. The preceding discussion has begun that job.

1.2.2d. Empirical Inadequacy

Empirical evidence for SSM is lacking, both in quality and quantity, in Miller’s initial presentation of the view. The weakest part of Miller’s empirical case for SSM consists of appeals to anecdote and supposed common knowledge. For example, take his claims about philanthropic millionaires’ pillow talk with their mistresses, or about waitresses’ perceptions of men’s tipping habits (2000: 323, 326). Such claims are not implausible but neither are they solid support for SSM.

Stronger but still shaky as support for SSM are Miller’s observations of modern charitable giving. Recall, Miller emphasises two features typical of modern charities: low efficiency (in terms of benefits to the needy conferred for donor costs incurred) and high concern for publicity. He suggests that, in
light of these features, giving to charity is best understood as costly signalling. It may well be, however, that the features of charity Miller emphasises can be explained without appeal to costly signalling. Perhaps charities spend big on publicity to allay potential donors’ sensible wariness of fraud and, as a result, operate inefficiently in the terms laid out above.

The most substantial planks from which Miller constructs his empirical case for SSM are the show-off hypothesis about big-game hunting and Buss’ cross-cultural mate preference data. As I will explain in what follows, though, both the show-off hypothesis and Buss’ data have serious flaws as support for SSM.32

The show-off hypothesis is meant to lend plausibility to SSM but that hypothesis is itself controversial. There are many theories about hunting and food sharing. There is also much data on patterns of hunting and food sharing across cultures. I will draw here on Michael Gurven’s (2004) evaluation of the major theories in relation to an impressively detailed review of the data.

According to kin selection-based nepotism (KS), hunters’ “food sharing should favour biased transfers toward kin” (2004: 545). According to reciprocal altruism (RA) hunters will “give portions of food to individuals with whom one has shared in the past, and from whom one is likely to receive shares in the future” (2004: 545). According to the tolerated theft (TT) view, hunters “should cede portions to other individuals if [the] price of defence is greater than the additional value that could be gained from consuming those extra pieces” (2004: 545). Finally, on costly signalling (CS) accounts (of which the show-off hypothesis is one), the distribution of hunting spoils is expected to be extensive “because the payoffs to signalling derive only from the honest display of production to a wide audience, and not from giving to specific individuals” (2004: 546).

32 There is a rich vein of empirical work relevant to SSM running through the fields of psychology, economics and anthropology. My criticisms here are meant to motivate my exploration of that work in Chapter 5.
In order for KS and RA to be plausible, "producer control" must be a feature of big game hunting and sharing (Gurven [2004: 546]). That is, big game hunters must be able to influence who gets how much of the meat they acquire. This need not be the case in order for TT or CS to be plausible. In order for RA to be plausible, not only producer control but also "contingency" must be a feature of big game hunting and sharing (Gurven [2004: 546]). That is, recipients of hunters' largesse must provide meat (or some other resource) in return. This need not be the case in order for KS, TT or CS to be plausible. Recall, Hawkes' (1991) show-off hypothesis was motivated in part by the claim that producer control and contingency are largely absent in the sharing of big game. If this were true, then CS (along with TT) would enjoy a substantial boost in plausibility over KS and RA.

Cross-cultural data on hunting and food transfers suggests that producer control and contingency are not largely absent in the sharing of big game (Gurven [2004: 548-549, 550-552]). CS does not stand out above KS and RA in light of the available data on food sharing. On the contrary, Gurven writes:

> Although men's focus on game production may be motivated, in part, by the mating benefits of signalling, hunting seems to be a viable provisioning strategy...and does not require costly signalling to justify its widespread occurrence (2004: 559).

The supposed lack of producer control and contingency in the sharing of big game hunting spoils were not the only things Hawkes took to favour explaining big game hunting in terms of signalling. The supposed inefficiency of hunting big game, as compared to hunting smaller game or gathering, was also important in supporting the show-off hypothesis. Hawkes claimed that male big game hunters of the Paraguayan Ache tribe:

---

33 Gurven synthesizes results from a great array of studies of numerous cultures. I will not attempt to report on those studies here. I will, however, emphasize a caution Gurven offers about reading the cross-cultural record. He warns against making "implicit assumptions that lack of control is signified by a hunter's receiving 1/n, and that complete control is viewed as an ability to hoard 100% of a resource" on the grounds that "keeping 1/n does not signify a lack of control if the acquirer decides that 1/n is the optimal portion to keep, given the expected payoffs from sharing" (2004: 548). Such assumptions might underlie the claim that data on hunting and food sharing supports CS over rival theories, especially KS and RA.
...sometimes preferentially target resources with expected energetic return rates which are lower than available alternatives [and] could supply more calories to their families for the same time spent if they chose otherwise (1991: 30).

The available alternative for Ache hunters is, supposedly, palm starch. Drawing on work by Hill et al. (1987), Hawkes claimed that hunters could roughly double their mean rate of return (measured in calories) by gathering palm starch instead of pursuing big game. If this were correct, then Hawkes’ show-off hypothesis (or CS accounts more generally) would seem appealing, insofar as it identifies a goal other than provisioning that makes sense of what big game hunters are doing.

Whether big game hunting really is inefficient is questionable. Methodological flaws in the work by Hill et al. (1987) have recently been discovered, and new data on foraging return rates among the Ache, the Hazda of Tanzania and the Venezuelan Hiwi suggests that hunting big game produces rates of caloric return higher than gathering or hunting small game (Gurven & Hill [2009: 52-53]). It is thus far from clear that big game hunting is such a sub-optimal provisioning strategy as to push us toward a costly signalling explanation for the practice.

The idea that the inefficiency of big game hunting would automatically lend credence to a CS-type account of that activity is also questionable. Even if big game hunting really is inefficient as compared to, say, gathering, there are non-CS-type explanations available, purely in terms of provisioning, for why men would hunt anyway:

[A] sexual division of labour is expected when multiple currencies (e.g., protein-lipid, carbohydrate) provide utility, when the activities that produce them are mutually exclusive, when either sex has a comparative advantage, and when high productivity requires a relatively long training period (Gurven [2004: 554]).

Even if men’s hunting is inefficient taken in isolation from women’s foraging, men’s and women’s patterns of hunting and foraging may together be the result of an adaptive sexual division of labour. Meat contains important nutrients not obtainable from plant products. Successful hunting requires
much practice. Men are less burdened by childcare and generally bigger and stronger than women. Finally, were both sexes to exploit the same food sources, the supply of those foods may well deplete. In light of all this, it may well be that men and women specialise in different skills and exploit different resources, pooling their hauls and thus increasing overall efficiency in obtaining food (Gurven & Hill [2009: 57]). Thus, data showing big game hunting to be less efficient than other approaches to obtaining food does not necessarily favoured the show-off hypothesis, or CS accounts more generally.

Miller presents Hawkes’ show-off hypothesis in order to make it plausible that appealing to signalling to account for puzzlingly costly and other-benefiting behaviours is, at least sometimes, a viable explanatory strategy. The show-off hypothesis is far from secure, though, and thus in no position to lend support to SSM. That is not to say a better case cannot be made for a CS-type account of hunting. Indeed, I will argue later (see §5.3.1) that turtle hunting among Meriam islanders, as described by Rebecca Bliege-Bird and her collaborators (e.g. Bliege-Bird et al. [2001]), is a plausible example of hunting as costly signalling. For now, my goal is only to raise problems with the empirical aspect of Miller’s initial case for SSM.

The next major line of evidence for SSM in Miller’s initial presentation of the view is research by Buss (1989) on mate preferences. As described earlier, Buss found that both males and females across 37 cultures considered the most desirable trait in a long-term mate to be kindness (1989: 13). Buss’ work supports the claim that moral behaviour is attractive to potential mates but there is a problem with taking this work as support for SSM. Crucially, such work cannot tell between the claim that moral behaviour is a sexually-selected signal and the claim that moral behaviour is merely a cue in mate choice, that is, a good thing to look for in a potential mate but not itself a signal. Were moral behavior merely a cue, sexual selection and in particular the power of mate choice to shape human traits could remain important elements in the evolution of morality (as Miller claims) but signaling theory and in particular the handicap principle would drop out of the picture.
Suppose we were to conduct a massive cross-cultural survey of mate preferences asking about the importance, not of kindness, wealth and beauty in a potential mate, but rather of sight and hearing. I confidently predict that both sexes, cross-culturally, would show a strong preference for sighted, hearing-intact mates. Supposing we found just that, we would not thereby by led to suppose that eyes and ears evolved as signals favoured in mate choice. Similarly, demonstrating the existence of a preference for kind (or generous, or honest) mates does not establish that those features evolved as sexually-selected signals. While this point may sound obvious, it is missed in Miller’s initial presentation of SSM.

The conceptual clarifications offered earlier – distinguishing signals from cues and highlighting complexities in the notion of cost – are useful now in spelling out just what it takes to empirically substantiate a costly signalling hypothesis. Showing that some trait $X$ affects the behaviour of others (or at least figures importantly in their decision-making) is only part of the empirical burden incurred by claiming $X$ is a costly signal. To empirically substantiate a costly signalling hypothesis about $X$, it must be shown that:\textsuperscript{34}

(1) receivers respond differentially to $X$
(2) $X$ is an indicator of qualities of interest to receivers
(3) $X$ is a reliable indicator of such qualities
   (a) $X$ is costly in terms of some component of fitness
   (b) $X$ is differentially costly for honest vs. dishonest senders

Recall Maynard-Smith & Harper’s first condition for some act or structure qualifying as a signal, that is, the act or structure must alter the behaviour of some other organisms. Requirement (1) above essentially restates this condition. Recall also Maynard-Smith & Harper’s third condition, that is, other organisms’ responses to the act or structure must be evolved responses. Meeting requirement (2) is a matter of showing a link between $X$ and something else in virtue of which certain other organisms benefit from attending to $X$.\textsuperscript{35} Finally, meeting requirements (3a) and (3b) is necessary to

\textsuperscript{34} I draw here on Stewart Saunders’ (2009) review of Searcy & Nowicki’s (2005) book on costly signalling.
\textsuperscript{35} The paradigmatic example here is the link between the features of a peacock’s tail – the signal – and his genetically-based immunocompetence, which is a highly fitness-relevant
establish that $X$ is a reliable signal *because* it is costly and not because, for example, it is impossible to fake, as in case of an index. Applying these requirements to the particular case of SSM yields the following set of questions:

(1) Do potential mates respond differentially to displays of moral behaviour?
(2) Does moral behaviour indicate qualities of interest to potential mates?
(3) Is moral behaviour a reliable indicator of such qualities?
   (a) Is moral behaviour costly in terms of some component of fitness?
   (b) Is moral behaviour differentially costly for honest vs. dishonest senders?

As is probably apparent, Buss' data goes only a short way toward empirically establishing SSM. At most, that data gestures at a positive answer to (2) and even then only on the (substantial) assumption that behaving morally correlates well with being kind. A great deal more in the way of empirical evidence is needed to substantiate SSM as a hypothesis about the evolution of morality.

1.2.2e. Philosophical Implications

In Miller's initial presentation of SSM, the worry that SSM threatens to cheapen or debase morality is dismissed on the grounds that to ascribe a sexual function to moral behaviour is not to ascribe a sexual motivation to moral agents. Miller is right; we should set aside that naive worry. Miller's treatment of the broader implications of SSM is fine as far as it goes, but it does not go nearly far enough.

Recall Kitcher's framework for thinking about the ways in which biology might bear on ethics, that is, by revealing:

(1) how people come to acquire ethical concepts, to make ethical judgements about themselves and others, and to formulate systems of ethical principles
(2) facts about human beings that, in conjunction with moral principles that we already accept, can be used to derive normative principles that we had not yet appreciated
(3) what ethics is all about and [so] settle traditional questions about the objectivity of ethics
(4) new normative principles

If possibilities (2) and/or (4) were realised, then biology would have implications for normative ethics, where this is understood to be, roughly, philosophising that aims to determine what we morally ought to do. If possibility (3) were realised, then biology would have implications for metaethics, where this is understood to be, roughly, philosophising that investigates the nature of moral discourse, the existence and nature of moral properties, and the epistemological status (truth or justification) of moral claims.

One way to pursue (3) is by fleshing out (1). The thought is that an explanation for why we make moral judgements may license conclusions about the truth or justificatory status of moral judgements and/or about the existence of moral properties. I will take as my departure point here the work of Michael Ruse & Edward Wilson (1986), who made an early attempt to draw metaethical conclusions out of an evolutionary account of morality using just this strategy. Their attempt was much-criticised and ultimately unsuccessful but illuminatingly so.

Ruse & Wilson offered an evolutionary explanation of morality on which the function of thinking in moral terms is, broadly speaking, to promote fitness-enhancing cooperation. They took the upshot of their account to be that "our sense of 'right' and the corresponding sense of 'wrong' [are] brought about by ultimately biological processes" (1986: 179). Part of what our moral sense tells us, according to Ruse & Wilson, is that some things are right and others wrong independent of what we humans happen to believe or desire. Ruse & Wilson's description of moral phenomenology is not idiosyncratic. In a mid-20th century treatment of moral phenomenology, philosopher Maurice Mandelbaum wrote:
It is my contention that the demands which we experience when we make a direct moral judgement [i.e. a judgement of rightness or wrongness made when in a situation requiring a moral choice] are always experienced as emanating from "outside" us, and as being directed against us. They are demands which seem to be independent of us and to which we feel that we ought to respond (1955: 54).

More recently, Terry Horgan & Mark Timmons have noted the many references in metaethical debate to moral phenomenology's "objective purport" ([2008: 267]; see also their [2005]). That term is Jonathan Dancy's (1986: 239) but similar claims can be found in the work of Mackie (1977), Brink (1989), and Huemer (2005). Whether our moral phenomenology is univocally objectivist is open to question (Loeb [2007]), but Ruse & Wilson are on fairly solid ground at this step in their argument.

Ruse & Wilson claim that our sense that morality is objective is misleading. We have this sense because "human beings function better if they are deceived by their genes into thinking that there is a disinterested objective morality binding upon them" (1998: 179). As Ruse put things elsewhere:

[M]orality simply does not work...unless we believe that it is objective. Darwinian theory shows that...morality is a function of (subjective) feelings; but it shows also that we have (and must have) the illusion of objectivity' (1998: 253).

Kitcher was initially intensely critical of this kind of approach to pursuing possibility (3). Beliefs can be true despite dubious origins and can become justified even if initially unjustified, thus, we are not entitled "to infer that a natural history of ethical behaviour will automatically show that our ethical beliefs are not objectively justifiable" (Kitcher [1985: 419]). Elliott Sober, another critic of early Darwinian metaethics, agreed with Kitcher about the lack of an "automatic connection" between explaining a judgement and justifying that judgement (1994: 94).

Kitcher and Sober were right to criticise Ruse & Wilson and to caution others interested in pursuing (3) via (1). Identifying the origin of moral judgements does not automatically generate any conclusions about the truth or justification of moral judgements. This does not, however, mean there are no
moves to be made from an explanation of moral judgements to a metaethical conclusion. Indeed, while Kitcher was initially hostile to attempts to pursue (3) via (1), he was later more open-minded about this project:

[T]he connection between Darwinian reforms in metaethics [i.e. (3)] and tracing the history of our moral attitudes [i.e. (1)] is more intimate than we might have thought. We can't simply assume that a historical investigation will leave everything in place. ([2003: 413]; see also [2005])

Empirically substantiating (1) might have a metaethical upshot, but identifying it requires more finesse than Ruse & Wilson display. Numerous philosophers have attempted the task since Ruse & Wilson's early effort. Some of them have been of the view that the metaethical upshot of an empirically substantiated evolutionary story about (1) is, by and large, positive: telling us that moral properties really do exist (and something about their nature), for instance, or that at least some moral claims are true or justified. I will briefly detail several examples of such views in what follows.

Robert Richards (1986) and Richmond Campbell (1996) both take evolutionary biology to justify particular moral claims. Richards thinks an appeal to evolutionary biology can justify the moral principle "foster the community good" (1986: 281). Campbell seeks to similarly justify the claim that human beings ought to follow some moral code or other rather than none at all (1996: 24). Larry Arnhart (1998) and William Casebeer (2003) both attempt to recast Aristotelian virtue ethics in a scientifically-respectable form, using an understanding of functions and natural kinds borrowed from evolutionary theory. Casebeer says that "moral claims should be reduced to functional claims technically construed [i.e. as construed in evolutionary biology]" and that "on this picture, moral facts are not 'queer' and unscientific" (2003: 4). Arnhart identifies "twenty desires that are universal because they are part of the biological nature of human beings" and claims that "these natural desires support universal standards of moral judgement" (1998: 17). William Harms marries a speculative adaptive history of moralising to a teleosemantic theory of truth on which the "truth of a signal just is the fact that the signal now stands in the kind of relation to the world
which explains the historical contribution to fitness of the signalling system” (2000: 701). The offspring of this union is the view that our moral judgements are true often enough to consider moral discourse largely successful rather than riddled with errors. Finally, William Rottschaefer & David Martinsen (1990) give an account of moral properties, such as goodness and rightness, by drawing on an evolutionary understanding of morality. Rottschaefer & Martinsen contend as “robust Darwinian naturalists” that:

...moral properties are biologically emergent relational properties... constituted by the relational triad of moral sentiments, as evolutionary adaptations, the objects to which they incline us, and the fitness that their achievement entails (1990: 161, 163).

Richard Joyce (2006: 143) usefully divides folk who think (1) has a metaethical upshot into those who take that upshot to be vindicating and those who take it to be undermining of morality. The views just surveyed are of the former kind. On the latter kind of view, by contrast, an evolutionary story about (1) in some way counts against the existence of moral properties or against the claim that any moral judgements are true or justified. Ruse & Wilson’s view that the objectivity of morality is an evolution-induced illusion falls on the ‘undermining’ side of Joyce’s divide. So does Joyce’s own view that empirically establishing that our tendency to make moral judgements is an adaptation “amounts to the discovery that our moral beliefs are the product of a process that is entirely independent of their truth” and so “should undermine our confidence in moral thinking” (2008: 218). Joshua Greene also counts among the underminers. Greene suggests that biology, especially evolutionary biology and neuroscience, provides “the beginnings of a debunking explanation of moral realism”, which he understands as the view that “there are genuine moral facts” (2003: 849). Finally, Sharon Street argues that “realist theories of value”, according to which “at least some evaluative facts or truths...hold independently of all our evaluative attitudes”, are “unable to accommodate the fact that Darwinian forces have deeply influenced the content of human values” (2006: 109, 110).

---

36 See Ruth Millikan (1984) for more on teleosemantics.
The fine details of attempted Darwinian debunkings of morality need not detain us at this moment (such details will be covered in detail in §6.1). The goal for now is only to emphasise that it is a matter of much dispute just what implications an evolutionary explanation for morality would have: vindicating, undermining, or none at all? It is safe to say that the implications (if any) of an evolutionary explanation for morality are highly dependent on the empirical details of such an explanation. That said, a real question arises about how SSM, if true, would bear on the dispute between vindicators and debunkers. Miller does not consider this question in his initial presentation of SSM. This is not a failing on his part; nobody can do everything. It does leave much unexplored intellectual territory, though, which I will enter in later discussion (see especially §6.4).

1.3. Summary and Preview

In this chapter I have presented Miller’s (2000) initial version of SSM, the hypothesis that sexual selection was an important factor in the evolution of human morality. As it stands, SSM is under-developed both conceptually and empirically. My criticisms in this chapter have identified several areas in which SSM needs clarification and elaboration. Nevertheless, Miller’s initial presentation of SSM does put on the table an interesting view about the evolution of morality that is worth taking seriously. In the chapters to come, I will attempt to develop SSM by showing how the conceptual confusions I have identified here can be clarified, by assembling supporting evidence drawn from several disciplines, and by carefully considering the possible metaethical implications of SSM. My general strategy in the chapters to come will be to present, evaluate, and extend Miller’s later (2007) work on SSM. As mentioned earlier, this later work partially addresses some of the criticisms I have offered in this chapter but still leaves much to do in the way of making SSM conceptually clear and empirically well-supported.

Chapter 2 takes up the issue of how to specify the explanatory target of an evolutionary account of morality. I will survey a range of work done under the rubric of ‘evolutionary explanation of morality.’ I suggest that accounts of
the evolution of morality can be divided into those which attempt to explain specific kinds of behaviour, those which attempt to explain certain social or group-level phenomena, and those which attempt to explain some element of human psychology. There is significant variation within each of these categories but it is a variant of the last kind of account, I will argue, that identifies the appropriate target of explanation.

My next goal in Chapter 2 will be to consider Miller’s updated (2007) version of SSM, which explicitly attempts to integrate SSM with virtue ethics. I claim that framing SSM in terms of virtue ethics still fails to get the explanatory target right. I go on to propose a fix for SSM. Essentially, the fix is to marry SSM to an account of the evolution of morality that does get the explanatory target right. The partner I have in mind for SSM is Joyce’s (2006) account, which seeks to explain the evolution of the tendency to make moral judgements by appealing to the motivational efficacy of such judgements.

Developing SSM in this way will require defending Joyce’s approach to explaining the evolution of morality. Stephen Stich has been a particularly prominent critic of Joyce’s approach (Stich [2008]; Kelly et al. [2007]; Kelly & Stich [2008]). Responding to Stich’s objections to Joyce’s approach will be the last task of Chapter 2.

In Chapter 3 my goal is to consider how sexual selection fits alongside other processes that many have taken to be important in the evolution of morality. The whiff of greedy theorising about Miller’s initial presentation of SSM has mostly dissipated in his later (2007) work; there he gives a slew of other processes their due. I will briefly describe how mutualism, kin selection, reciprocity, and group selection might contribute to an evolutionary explanation of morality. I will devote a great deal more attention to an explanation for our moral capacities that remains overlooked even in Miller’s later discussion of SSM: the by-product explanation. Jesse Prinz has recently championed this view, arguing that our capacity to make moral judgements is “an accident”, the “by-product of capacities that were evolved for other purposes” ([2009: 168]; see also [2007a]; [2008a]; [2008b]). If Prinz were
correct, this would be bad news for adaptationist accounts of morality, or which SSM is one. I will argue against Prinz, defending the viability of views on which the making of moral judgements is given an adaptive explanation.

The remainder of Chapter 3 will be given over to assessing Miller’s claim (made in his initial work on SSM and maintained in his later work) that sexual selection is a special kind of evolutionary process distinct from natural selection. I find Miller’s arguments for this claim unconvincing. I next draw on the biological literature to find or reconstruct arguments for the distinctness of sexual and natural selection. Numerous biologists have expressed views about the relationship between sexual and natural selection. While there is considerable disagreement, though, there is precious little debate on this issue. Chapter 3 is my attempt to see what (if anything) of substance hangs on whether sexual and natural selection are distinct evolutionary processes. I tentatively conclude that there is probably not a sustainable distinction to be drawn between natural and sexual selection. Finally, I show how this conclusion need not vitiate SSM, which can remain an interestingly novel view about the evolution of morality even if sexual and natural selection are not distinct.

Chapter 4 continues the project of clarifying SSM. I will use the distinctions introduced earlier in this chapter – between signals and cues, expenditures and costs, absolute costs and relative costs, and handicaps and indicies – to work carefully through Miller’s ([2000]; [2007]) proposals regarding morality, signalling and inter-sexual selection. I will also distinguish importantly different versions of the costly signalling idea, which has gone through several incarnations since its inception in the work of Zahavi (1975). Searcy & Nowicki (2005: 10) distinguish “Zahavi handicaps” from “conditional” and “revealing” costly signals. Lachmann et al. (2001: 13189-90) think the key insight in the costly signalling idea applies even when honest signals are cheap or free, so long as dishonest signalling is prohibitively
I claim that getting clear on the distinctions drawn above and on the differences between the various versions of costly signalling theory allows for a more sophisticated and accurate understanding of the ways in which moral behaviour and judgement may serve as reliable signals.

Chapter 5 will evaluate the empirical evidence for SSM. I will begin by clarifying the empirical commitments of SSM, in the hope of forestalling misguided objections. Specifically, I will clarify SSM’s predictions regarding the existence and extent of sexual dimorphism in morality, and the development and expression of moral capacities across life-history stages.

After clarifying the empirical commitments of SSM, I will survey a range of studies conducted in psychology, anthropology, and experimental economics. The results I assemble support SSM insofar as they support the idea that moral behaviour plays a signalling role. However, these results suggest that SSM is importantly incomplete. SSM faces what I will call the many receivers problem. The signalling role of morality is not limited to attracting potential mates but rather extends to obtaining cooperative partners more generally.

My final goal in Chapter 5 will be to highlight the contribution that sexual selection in particular can make to a more general signalling-based account of morality. Sexual selection is, I claim, vital in understanding moral behaviours that serve as signals of ‘good genes.’

Finally, Chapter 6 will consider the philosophical implications of the truth of SSM. I will briefly consider Miller’s (2007) discussion of SSM’s implications for moral philosophy before moving on to the details of attempted Darwinian debunkings of morality. I will identify two different

---

37 According to Lachmann et al., the key insight is that if signalling is “differentially costly according to the meaning and/or validity of the message [then] signallers can be relied upon to report aspects of their condition or of the environment” (2001: 13190). The crucial thing on their view is the relative cost of honest versus dishonest signals rather than the absolute cost of honest signals (as was the case in Zahavi’s first formulation of the Handicap Principle).
debunking strategies, namely, metaphysical debunking and epistemological debunking.

The remainder of Chapter 6 will focus on the epistemological debunking strategy and, in particular, on a prominent line of reply to this strategy, which I call the reliabilist reply. I will lay out the conditions that must be met for that reply to succeed. If SSM is true, it will be argued, then not all the conditions for the success of the reliabilist reply are met. In particular, what I will call the tracking condition fails to be met. The truth of SSM, I claim, counts against the plausibility of the reliabilist reply to the Darwinian debunkers.
Chapter 2. The Evolution of Morality: Targets of Explanation

2.1. Introduction

This chapter takes up the issue, raised in Chapter 1 (see §1.2.2a), of how to specify the explanatory target of an evolutionary account of morality. My first task will be to set aside several things that are not the target of such an account. Next, I will survey a range of work done under the rubric of ‘evolutionary explanation of morality.’ Accounts of the evolution of morality, I claim, can be divided into those which attempt to explain specific kinds of behaviour, those which attempt to explain certain social or group-level phenomena, and those which attempt to explain some element of human psychology. It is a variant of the last kind of account, I will argue, that identifies the appropriate target of explanation. While Miller’s updated (2007) version of SSM falls into the latter category, I claim this version of SSM still fails to get the explanatory target right. I will propose a fix for SSM, namely, to conjoin it with an approach to the evolution of morality that does get the target right. Specifically, I will show how SSM can be merged with the approach taken by Richard Joyce (2006), who seeks to explain the evolution of the tendency to make moral judgements.

Developing SSM in this way will require defending Joyce’s approach to explaining the evolution of morality. Stephen Stich is a particularly prominent critic of Joyce’s approach (see e.g. Stich [2008]; Kelly et al. [2007]; Kelly & Stich [2008]). Stich questions whether moral judgements as understood by Joyce are human universals. Stich also suggests that Joyce’s focus on helping, harming and fairness in explaining the evolution of morality is overly narrow. I will present and respond to Stich’s concerns. In response to the first concern, my strategy will be to undermine a key piece of empirical work (Kelly et al. [2007]) upon which Stich relies. In response to the second concern, I will show how the observations Stich makes about the extent of the moral domain are not incompatible with the kind of evolutionary explanation of morality Joyce offers.
2.1.1. Untangling Terminology

Terminological confusion plagues discussions of the evolution of morality. The first task here will be to set aside several things that are not the target in attempting to give an evolutionary explanation of morality. Doing so will require discussing altruism, selfishness and cooperation in the context of evolutionary theory. This task is tedious but necessary: confusions continue to crop up, not just in popular forums but even in the recent work of specialists.\(^{38}\) A provocative and much-quoted passage from biologist Michael Ghiselin provides a good entrée into this discussion:

> Where it is in his own interest, every organism may reasonably be expected to aid his fellows. Where he has no alternative, he submits to the yoke of communal servitude. Yet given a full chance to act in his own interest, nothing but expediency will restrain him from brutalising, from maiming, from murdering his brother, his mate, his parent, or his child. Scratch an “altruist” and watch a “hypocrite” bleed (1974: 247).

Evolved agents, Ghiselin suggests, will act in the interests of others only when doing so is in their own interest. While expediency of this kind results in agents doing much that advances others’ interests, taking such agents to be truly altruistic would be naïve (and taking oneself to be so would be hypocritical or at least deluded). The suggestion here is that genuine altruism — and not just scare-quotes “altruism” — requires acting in others’ interests to the detriment of one’s own. That, Ghiselin suggests, we should not expect of evolved agents. As Ghiselin puts the idea elsewhere:

> [I]f the hypothesis of natural selection is both sufficient and true, it is impossible for a genuinely disinterested or altruistic behaviour pattern to evolve. Naturally, this fact suggests that many traditional notions about ethics are wrong (1973: 967).

Ghiselin’s discussion of evolution, altruism and ethics is densely packed with confusions. Digging these out and clarifying them is worthwhile exercise in working toward the goal of identifying the appropriate explanatory

---

\(^{38}\) The New York Times Science News regularly features articles on evolution and morality (usually riffing off the work of a high-profile scientist who has a new book out) that with typical journalistic blitheness run together many of the notions I will go on to distinguish in this section. Allen & Bekoff’s (2005) discussion of “wild justice” is an example of academic work on the evolution of morality that is similarly frustrating.
target for evolutionary accounts of morality. Ghiselin jumbles up the consequences of acts with the motives for acting, agents’ self-interest with their fitness, and talk about altruism with talk of ethics.

The first confusion to be addressed is between the consequences of and the motivations for actions. Ghiselin claims that natural selection will produce agents that act only in ways that advance their own interests, which is a claim about the consequences of acts. Even if Ghiselin’s claim is correct, his charge that any supposed altruism is in fact hypocrisy sticks only if evolutionary theory tells us something about agents’ motives for acting. Nothing about an agent’s motives for acting in a certain way follows directly from an observation of the consequences of the agent’s action. Absent a story about how to draw a conclusion about motives from a claim about consequences, Ghiselin’s charge is unsubstantiated.

The second confusion that needs highlighting is between self-interest and reproductive fitness. Ghiselin claims that natural selection militates against disinterested behaviour and only favours behaviour that advances agents’ own interests. This makes sense only if he is equating agents’ interests with their reproductive fitness. Doing so, however, is a mistake. Doing so ignores the important distinction between individual organisms and genes or, as Richard Dawkins (1989/1976: 254) put it, between the “vehicles” in evolution and the “replicators.” What is good for an individual agent and what will get copies of that agent’s genes into future generations can come apart, and quite radically. Take for example any case in which an organism can ensure its genes make it into the next generation only by sacrificing its own life. So, even if Ghiselin is right that natural selection produces agents that act always so as to advance their reproductive fitness, he is not therefore entitled to conclude that evolved agents only ever act in their own interests.

Finally, Ghiselin moves swiftly from talk about evolution and altruism to talk about ethics via the easy assumption that being moral simply is (or at

---

39 Ghiselin’s discussion is a goldmine of sorts but, given that what it yields is mostly confusions, the better metaphor would probably be pyrite mine.
least requires) being disinterested. Whatever morality really is or is really all about (a question that will be considered in detail later in this chapter), Ghiselin’s view on the matter is inadequate. Altruism and morality cannot be breezily equated. Promise-keeping provides a counter-example to the claim that being moral is a matter of acting in ways that advance others’ interests at a cost to one’s own, since the promise keeper’s own interests are often advanced. Cases of nepotism and other sorts of unfair discrimination motivated by a desire to advance the interests of the favoured individual provide counter-examples to the claim that being moral is a matter of being motivated to advance others’ interests. As in the case of self-interest and reproductive fitness, then, morality and disinterestedness can come apart.

Much of the confusion in Ghiselin’s discussion can be traced to an unresolved ambiguity in the term ‘altruism.’ Elliott Sober (1993a) did much to resolve the ambiguity by distinguishing between a vernacular and a technical, evolutionary biologist’s sense of ‘altruism.’ While certainly an advance on Ghiselin’s discussion, Sober’s treatment does leave something to be desired. In what follows, I will present Sober’s view and what I take to be its shortcomings. My overall aim remains clarificatory: to describe the various explanatory targets that pop up in discussions of human evolution and to recognise those that are, for my purposes, distractions.

Sober defines ‘evolutionary altruism’ as follows:

A trait is altruistic in the evolutionary sense if two conditions are satisfied. First, bearers of the trait are less fit than nonbearers of the trait within the same group. Second, groups of altruists have a higher average fitness than groups of selfish individuals. (1993a: 205)

The reference to groups in Sober’s definition is important in the context of Sober’s own views on the possibility of evolutionary altruism (see also Sober & Wilson [1998]). Whether Sober’s definition actually captures
what biologists mean by altruism, though, could be questioned.\textsuperscript{40} For instance, Robert Trivers, a key figure in modern evolutionary biology, writes:

\[\text{Altruistic behaviour can be defined as behaviour that benefits another organism, not closely related, while being apparently detrimental to the organism performing the behaviour, benefit and detriment being defined in terms of contribution to inclusive fitness (1971: 35).}\]

No reference to groups appears here. Nor does such a reference appear in the definitions of altruism offered by other prominent biologists (see e.g. Hamilton [1964: 15]; Wilson [2000: 121, 578]). For this reason, while recognising the importance of Sober’s clarificatory work, I prefer to use terminology drawn from Joyce, who captures evolutionary altruism with the term “fitness sacrificing,” which he defines as “behaving in a way that advances another individual’s reproductive fitness, at the expense of one’s own reproductive fitness” (2006: 13). Joyce’s term also has the advantage of avoiding any potential for confusion with the other sense of altruism, to be detailed below.

Returning to Sober’s distinction between two senses of altruism, ‘psychological altruism’ is defined by first defining two types of preference. A self-directed preference “describes what one wants for one’s self, but does not say anything about the welfare of others,” while an other-directed preference “describes what one wants for others, but does not say anything about what one wants for one’s self” (Sober [1993a: 207]). Psychological altruists, according to Sober, put the satisfaction of other-directed preferences ahead of that of self-directed preferences in at least some cases of conflict between the two. Psychological egoists, by contrast, never satisfy other-directed preferences at the expense of self-directed ones.\textsuperscript{41}

\textsuperscript{40} See Kerr \textit{et al.}(2004) for a useful survey of the different senses of ‘altruism’ in play in the evolutionary biology literature.

\textsuperscript{41} Sober aims to characterize different kinds of agents. The names Sober gives to kinds of agents should not be confused with the names for theories about agents or for those who accept such theories, e.g. Hobbes is often labelled an Egoist on the grounds that he accepted Egoism, the thesis that all voluntary human action is ultimately motivated by self-interest.
An example may help in understanding Sober’s view of psychological altruism and egoism. Imagine two agents A and B who both want to be rich (a self-directed preference) and who both want to reduce third-world hunger (an other-directed preference). Imagine both agents could reduce third-world hunger but only by donating some money that could instead have been profitably invested. Suppose A donates the money, foregoing the gain in wealth that investing it would have brought; A counts as psychologically altruistic by Sober’s lights. If B invests the money rather than donating it, then Sober would count B as a psychological egoist.\(^42\)

As above, we can ask if Sober’s definition actually captures the relevant sense of ‘altruism.’ Also as above, there is reason to doubt it. On Sober’s account, whether an agent is a psychological altruist is not a matter simply of whether she has desires to benefit others purely for those others’ sake. Suppose Jenny has many such desires and that she often acts to satisfy them. Jenny could still be an egoist, in Sober’s terms, if she never resolves conflicts between benefiting others and benefiting herself in favour of the former.

Sober’s understanding of psychological altruism is in tension with classical discussions of altruism, which have taken the crucial question to be whether people ever care about others for their own sake and not merely instrumentally. Thomas Hobbes, for example, denied that altruism thus understood is a part of human psychology, claiming that:

\[
\text{...no man giveth but with intention of good to himself; because gift is voluntary; and of all voluntary acts, the object is to every man his own good (1982/1651: 209).}
\]

Joseph Butler was concerned to refute Hobbes and show how benevolence, meaning “an affection to the good of our fellow creatures,” was distinct from and compatible with self-love, “a regard to our own private

\(^42\) Sober further distinguishes two kinds of psychological altruist. Moderate psychological altruists satisfy other- rather than self-directed preferences in some but not all cases of conflict between the two, while Extreme psychological altruists always satisfy other- rather than self-directed preferences when the two come into conflict (1993: 208).
good” (1983/1726: 49). More recently, psychologist C. Daniel Batson has tried to show experimentally that some people, some times, experience “motivation with an ultimate [i.e. not merely instrumental] goal of benefiting someone else” and thus to show that “altruistic concern for the welfare of others is within the human repertoire” (1991: 2, 174).

That common usage would refuse to count an agent like Jenny above as an altruist seems at least questionable. Sober may have built more into his notion of psychological altruism than either philosophers or the folk require. Without pretending to have settled anything here, I will prefer to use terminology drawn from Joyce, who uses ‘altruism’ to mean “acting with the intention of benefiting another individual, where this is motivated by a non-instrumental concern for his/her welfare” (2006: 14).

To recap: the temptation to think that if we humans are evolved creatures then our being altruistic is unlikely if not impossible can be traced to a confusion between the technical and vernacular senses of ‘altruism.’ Once those senses are distinguished, it is clear that altruism and fitness-sacrificing can come apart. An altruist need not be a fitness-sacrificer. Nor must a fitness-sacrificer be an altruist. Neither altruism nor fitness sacrificing should be mistaken for the target of an evolutionary explanation of morality. Such mistakes are made. For example, in a review of attempted evolutionary explanations of morality, Lee Cronk writes that:

The growing literature on evolution and morality concerns one central question: Why do we have morality? How could a concern for the welfare of others evolve if, as is currently thought to be the case, natural selection only rarely favours traits that benefit groups but not individuals? (1994: 82).

43 For example, consider a family caught in a life-threatening flood and imagine the father refusing to take a seat in the overcrowded evacuation chopper so that his two children can fit in instead (and doing so not to avoid shame or guilt but out of a genuine desire that his children be safe).

44 The soldier who throws himself on the grenade that lands amidst his squad, the monk who swears (and keeps!) a vow of chastity, and the surrogate mother who sacrifices some of her childbearing potential to carry an unrelated couple’s child, can all plausibly be taken to advance others’ reproductive fitness at a cost to their own. None, though, must be altruistic. Perhaps the soldier only wanted to escape the horrors of war, the monk cares nothing for this world and merely seeks to secure the best possible spot in the next, and the surrogate mother is just in it for the money.
This is one question only if morality, altruism and fitness sacrificing are all run together. I have already said why equating altruism and morality is a mistake. One would hope examples are unnecessary to make the point that fitness sacrificing and morality should not be equated either.

One more thing needs to be clarified before moving on to discuss various attempted evolutionary explanations of morality. Cooperation in the natural world is sometimes presented as a puzzle for evolutionary theory. The behaviours discussed under the heading of ‘cooperation’ typically include food sharing, alarm calling, collective defence, and allparenting (see Dugatkin [1997] for a hefty collection of examples). In a pair of recent review articles, West et al. take the “problem of cooperation” in evolutionary biology to consist in explaining why organisms would engage in “behaviour that appears costly to perform but benefits other individuals” ([2007a: 415]; see also [2007b: 661]). Thus described, cooperation seems to be just fitness-sacrificing by a different name. More perspicaciously, one might say the puzzling phenomenon picked out by the term ‘cooperation’ is apparent fitness sacrificing, since closer examination tends to reveal a fitness benefit to the acting organism (e.g. the food sharer or the alarm caller).45

To appreciate what is meant to be distinctively puzzling about cooperation (as opposed to fitness sacrificing), it is perhaps necessary to don a dunce cap of sorts when it comes to biology. Tennyson’s line “Nature, red in tooth and claw” was biologically ill-informed but evocatively conveyed the idea that natural selection promotes ruthless, dog-eat-dog individualism.46 To one in the grip of that idea, organisms doing anything to help others will appear strange. I will thus eschew use of ‘cooperation’ in favour of Joyce’s

45 David Lahti (2003: 641) suggests that cooperative behaviours like those listed above be termed “ostensible altruism,” meaning cases in which it seems to us as if the agent incurs a net fitness cost in providing a benefit to others and we are (currently) ignorant of whether or not that is actually how matters stand.

46 Tennyson’s poem In Memoriam contains several verses reflecting on the patterns and processes apparent in Nature. The famous line quoted above sets up an opposition between observations of the natural world and belief in a loving God who orders and guides that world. Nature’s callousness features elsewhere in the poem too, when Tennyson says of her “So careful of the type she seems / So careless of the single life.” Notice, replacing ‘type’ with ‘gene’ would eliminate the naive group selectionist sentiment and produce some pleasing assonance to boot!
term "helping," defined as "behaving in a way that benefits another individual" (2006: 13).

As Joyce points out, helpful behaviour need not advance the fitness of its beneficiaries. However, the helpful behaviours prominently discussed in evolutionary theory all very plausibly have a positive effect on the beneficiaries' reproductive fitness: food sharing, alarm calling, collective defence and alloparenting are all helpful behaviours of this kind. Helpfulness of this kind presents a *prima facie* puzzle for evolutionary theory. Faced with such helpfulness, the challenge is to show one of two things. The first option is to show that the behaviour is not fitness-sacrificing for the helper. The second option is to show that, while the behaviour is fitness-sacrificing for the helper, this is nevertheless consistent with evolutionary theory.

As will be detailed in the next chapter (see §3.1.1), there are several paths by which evolution might have led to the existence organisms that help others. The goal for now is just to mark off helpfulness from fitness sacrificing and from altruism. Helpful behaviour need not be fitness sacrificing, nor must it be altruistic. Imagine, for example, a doctor who agrees to help a busy, professional, pregnant woman get rid of a healthy but very inconvenient baby and does so purely out of a desire to thumb his nose at the law.

Fitness sacrificing, altruism and helpfulness all crop up as targets of explanation in discussions of human evolution. I have been at pains to point out (and have perhaps belaboured the point) that neither of the first two things are the proper target of an evolutionary explanation of morality. Nor is the third. The example above shows, rather graphically, that whatever overlap there is between helpful acts and acts we consider morally required or laudable

---

47 See discussion of Ghiselin above, distinguishing the interests of an individual agent from that agent's fitness (the 'interests' of its genes).
48 By identifying a hidden fitness benefit to helping in that way.
49 This might be done by showing how fitness-sacrificing behaviour of that kind could evolve (e.g. by group selection). Alternatively, it might be done by showing that the particular instance of fitness-sacrificing helpfulness in question is simply a mistake, perhaps the result of the helper operating a novel environment or misperceiving the specific situation in which the helpfulness occurred.
(and that overlap may be considerable), the match is far from perfect. We can substitute ‘husband’ for ‘baby’ in the example above to make this point even more obvious (and dodge any controversy about the moral permissibility or otherwise of abortion).

Fitness sacrificing, helpfulness, and altruism are relevant when attempting an evolutionary explanation of morality. Indeed, helpfulness in particular will turn out to be crucially involved in the kind of evolutionary explanation of morality I favour. The point here is simply that fitness sacrificing, helpfulness, and altruism should not be confused with each other or with morality. That even the recent work of specialists sometimes betrays such confusions highlights the need for the clarificatory work above.

2.2. Explanations of Morality

What should evolutionary explanations of morality be explanations of? Those who offer evolutionary explanations of morality appear in many cases to be seeking explanations for quite different things, as will be made apparent by the discussion below. Of course, moral language is used to talk about many different things: right and wrong acts, good and bad rules, virtuous and vicious character traits, true and false moral beliefs, praise- and blameworthy people, and more besides. Given this diversity in moral discourse, the diversity of explanatory targets in discussions of the evolution of morality is perhaps unsurprising. Even so, it is worth asking what evolutionary explanations ought to target.

My approach to answering this question is unapologetically piecemeal. I consider it unfruitful to try to begin with a definition of morality and compare numerous and disparate so-called evolutionary explanations of morality to that standard. What morality is, and indeed how to go about defining it, are longstanding and contentious issues. Rather than try to settle them at the outset, I will instead survey a range of purported evolutionary explanations.

---
50 See Wallace & Walker (1970) for a collection of papers by philosophers discussing these issues, and Nado et al. (2009) for a more recent and empirically-informed treatment.
explanations for morality. Accounts of the evolution of morality, I claim, can be divided into those which attempt to explain specific kinds of behaviour, those which attempt to explain certain social or group-level phenomena, and those which attempt to explain some element of human psychology.\textsuperscript{51} In what follows, I will describe and discuss each kind of approach, along with specific examples of each. I will also ask what the explanatory target in each case actually is and what reasons there are for and against taking that to be the target at which an evolutionary explanation of morality should aim.

2.2.1. Explanations of Behavioural Morality

Some purported evolutionary explanations of morality are explanations for certain kinds of behaviour. Edward Wilson in \textit{Sociobiology} (2000)\textsuperscript{52} offered this kind of account. Wilson defined sociobiology as “the systematic study of the biological basis of all social behaviour” (2000: 4). Much of Wilson’s sociobiological work was devoted to understanding the behaviour of nonhuman animals. He did apply sociobiological thinking to our own species, though, offering explanations in terms of genetic fitness for why humans perform religious rituals, create art, wage war and, most importantly for current purposes, behave ethically (2000: 562-564).

Wilson’s view of what kinds of behaviour are ethical can be gleaned from a survey of his discussion and in particular the examples he discusses. Wilson appears to take behaving ethically to consist largely in helping others, reciprocating help, and punishing cheats (2000: 120-121, 129, 563). It should be clear, in light of the above discussion of helping, fitness sacrificing, and altruism, that Wilson’s approach to explaining the evolution of morality is deeply problematic. Wilson apparently makes the mistake warned against above, of equating morality with helpfulness, altruism, fitness sacrificing, or

\textsuperscript{51} I take this way of dividing the field to be natural and useful. Douglas Allchin (1999) offers a similar three-tiered framework as part of a guide to teaching about evolution and morality. He observes that research on the evolution of morality can focus on moral acts (the behavioural level), moral motivation or intent (the psychological level), or moral systems (the social level).

\textsuperscript{52} References are to the 25\textsuperscript{th} anniversary edition of Wilson’s \textit{Sociobiology}. The original was published in 1975.
some combination of the three. Recall also the distinction drawn in Chapter 1 between behaving in certain ways, on the one hand, and, on the other, judging that behaving in those ways is morally laudable or required. In light of this distinction, it should be clear that Wilson’s sociobiological story could at most explain why we evolved to behave (in some situations) helpfully or altruistically and not why we judge (some) helpful or altruistic behaviour to be morally required or laudable.

I take Wilson’s approach to explaining the evolution of morality to provide something of a cautionary tale. An evolutionary explanation of morality cannot simply target particular kinds of behaviour. Enough has been said, I think, to show why this is so.

2.2.2. Explanations of Social Morality

Some purported evolutionary explanations of morality are explanations for the prevalence across human societies of certain kinds of rule systems. A good example of this kind of approach to the evolution of morality is the work of Robert Boyd & Peter Richerson ([2002]; [2004]; [2005]). Boyd & Richerson’s key claim, for my purposes, is that “competition between [human cultural] groups will lead to the spread of moral systems that enhance group survival, welfare and expansion” (2005: 129, emphasis added). In understanding what is meant here by ‘moral systems,’ the key notion is that of a norm.

“Norms,” according to Boyd & Richerson, “are the result of shared notions of appropriate behaviour and the willingness of individuals to reward appropriate behaviour and punish inappropriate behaviour” (2005: 84). So, if the folk of group $G$ agree in condemning and punishing those who engage in behaviour $B1$, then $G$ has a norm forbidding $B1$. Or, if in $G$ those who do not engage in behaviour $B2$ are widely condemned and punished, then $G$ has a norm requiring $B2$. We can reasonably take moral systems to be sets of moral norms, but figuring out which norms are (according to Boyd & Richerson) the moral norms is less straightforward.
Boyd & Richerson include as moral norms “property rights” (i.e. rules against theft and vandalism) and also “rules against murder and assault” (2002: 287). These are examples of “group beneficial norms,” that is, norms which enhance the prospects of the group in competition with other groups. It would be a mistake to suppose, however, that Boyd & Richerson simply consider moral norms and group beneficial norms to be one and the same thing. Boyd & Richerson also count “the mutilation of genitalia and high rates of female infanticide” as moral norms, albeit “maladaptive” or “dysfunctional” ones unlikely to contribute to the success of groups that enshrine them (2005: 18).

One might be thus tempted to think moral norms on Boyd & Richerson’s account are any norms that govern interactions with others. This too would be wrong. In their view, “moral systems” also include “proscriptions against drunkenness, laziness, gluttony and other failures of self-control” (2005: 84). Moral norms can, it seems, cover purely self-regarding behaviours.\(^\text{53}\) As far as I can tell, nothing in Boyd & Richerson’s work on the evolution of moral systems makes clear just what distinguishes moral norms from non-moral norms, such as norms governing the wearing of hats (to use their own example). This is a problem, since the existence of such a distinction seems clear even if its precise details are not (see §2.3.1. below).

The same problem plagues related approaches to the evolution of morality. For example, Denis Krebs & Maria Janicki (2003) offer an account of the “biological foundations of moral norms” that owes much to Boyd & Richerson’s work. Krebs & Janicki at least offer an explicit view on what makes for a moral norm, writing that “all people acquire beliefs about how they should and should not behave [and] when such beliefs are adopted by most members of a culture, they constitute moral norms” (2003: 125). This is, however, manifestly inadequate. Most members of most cultures probably believe that one should not eat rotten meat or lick a stranger’s ear by way of

---

\(^{53}\) Of course, drunkenness and so on might be construed as social behaviours insofar as they can have effects (at times drastic) on others. Still, Boyd & Richerson seem willing to count norms against drunkenness and the like as moral norms even when the behaviour in question is “victimless” (2005: 84).
introducing oneself. Rules requiring abstention from eating rotten meat and ear-licking (at least with strangers) are not thereby moral rules. To treat them as such is to adopt an unnecessarily impoverished understanding of the kinds of rules (and rule violations) there are. Eating rotten meat is prudentially bad. Licking strangers' ears is a breach of etiquette. Neither is a moral matter, or at least, neither must be a moral matter simply because most people agree such things should not be done.

At risk of multiplying examples beyond necessity, I will mention one more case of a purported evolutionary explanation of morality that targets the prevalence of certain norms. William Harms & Brian Skyrms (2008: 434) aim to provide "an evolutionary explanation of the emergence of moral norms," by which they mean rules requiring certain kinds of behaviour, violations of which are met with social sanctions such as disapproval, punishment, and ostracism. Having stated their aim, Harms & Skyrms go on to discuss at length how populations embodying rules that require helpfulness or even fitness sacrificing might come to predominate over the course of evolution. As discussed above, though, morality, helpfulness, and fitness sacrificing come apart. It is a mistake to explain the evolution of norms requiring the latter two things and take oneself thereby to have provided an evolutionary explanation for distinctively moral norms.

The work by Harms & Skyrms, Boyd & Richerson, and others who take a similar approach to the evolution of morality may well represent significant progress in explaining human sociality. That is an interesting and important explanatory target in its own right. Whether such work explains human morality, though, seems to me dubious. An evolutionary explanation of morality cannot simply target norms (or rules, or systems) governing behaviour, without saying what is supposed to be distinctively moral about such norms (rules, etc.) and without saying why norms with those distinctive features (whatever they are supposed to be) would have evolved.

Finally, in spelling out what norms are, theorists who focus on the group level phenomena considered above seem to revert back (or down) to the
level of individual psychology. Norms are a matter of what people judge appropriate, are willing to punish, believe, and so on. Explanations for the evolution of morality that focus on social-level phenomena may thus bottom out in an account of some feature(s) of human psychology. I turn now to considering evolutionary explanations of morality that are conducted at that level of description.

2.2.3. Explanations of Psychological Morality

Some purported evolutionary explanations of morality are explanations for particular features of human psychology. This is by far the largest and most diverse class of purported evolutionary explanations of morality. I will begin this section by discussing a particularly prominent approach of this kind, which seeks to explain why we have as part of our psychological makeup certain emotions. The work of Robert Frank ([1988]; see also [2008]) provides a good example of this kind of approach.

Frank’s starting point is the idea of a “commitment problem” (1988: 4). A commitment problem arises whenever an agent can best serve his own interests only by credibly committing himself to act in a way that may later be contrary to his self-interest. An agent might need to make a credible promise of honesty in order to reassure and secure would-be partners in cooperative endeavours where cheating would be profitable and undetectable. An agent might need to make credible threats of revenge in order to deter would-be exploiters in situations where avenging a wrong would be more costly than not doing so. Commitment problems are common and solving them is important.

Frank suggests that evolution has endowed us with the means to solve commitment problems, namely, “moral sentiments” (1988: 46). Following Adam Smith, Frank counts as moral sentiments “anger, contempt, disgust, envy, greed, shame and guilt” (1988: 53). These emotions, Frank claims, help us solve commitment problems, since “being known to experience certain emotions enables us to make commitments that would otherwise not be credible” (1988: 5). The promises of an agent known to be prone to guilt will
be for that reason more trusted, Frank suggests, and threats from agents known
to be prone to anger will be for that reason taken more seriously.

Frank acknowledges that guilt- and anger-prone agents will sometimes
miss out on significant benefits (such as the benefits of undetected cheating) or
incur significant costs (such as the costs of exacting revenge). Overall, though,
the thought is that agents with an array of moral sentiments will do better (in
virtue of solving commitment problems) than will agents who lack such
emotions (and who cannot solve commitment problems). Moral sentiments,
then, are explained as the result of selective pressure on evolving agents to
come up with some way of solving commitment problems.

That Frank takes himself to be explaining morality seems clear.\(^\text{54}\) That
he is in fact doing that is less so. As was the case with Boyd & Richerson,
Frank’s work might well explain something important and interesting but it is
doubtful that thing deserves to be called morality.

The emotions Frank discusses could be considered moral sentiments in
the sense that they are emotions the having of which is judged to be morally
praiseworthy or blameworthy, good or bad, right or wrong, etc.\(^\text{55}\) Certainly,
moral judgements can be made regarding emotions. For example, it may be
judged that it is wrong (or at least bad) not to feel sympathy for innocent
victims of harm, that anger in response to unfairness is justified, that people
really ought to feel guilty if they cheat, and so on. Recall, though, there is
distinction between explaining the evolution of things (behaviours, emotions)
that are morally evaluated and explaining the tendency to make moral
evaluations (see §1.1.2a). Frank may be right that particular emotions have
evolved because they are effective (and reliably detectable) motivators and
inhibitors of action. To call Frank’s work an explanation of morality, however,
is overstating the case.

\(^{54}\) His book has chapters on moral sentiments, fairness, human decency, and “becoming
moral,” this last dedicated to the importance of emotions in the development of individuals’
moral competence.

\(^{55}\) I include the awkward ‘the having of which’ because it is not emotions themselves that we
morally evaluate but rather the agents who have them. We morally judge agents who get
angry, for example, not anger.
A couple of so-called fallacies are almost invariably raised in discussions of evolution and morality. The naturalistic fallacy is one, the genetic fallacy the other. I am tempted to introduce yet another to capture what goes wrong with many evolutionary explanations of morality: the synecdoche fallacy, which consists in taking an explanation for part of a thing to be an explanation for the whole thing. That emotions, helping, altruism, and systems of rules are importantly tied up with morality is undeniable, but an evolutionary explanation for any one of those things, however convincing, does not constitute an evolutionary explanation of morality.

2.2.3a. SSM and Virtue Ethics

Another kind of approach to the evolution of morality that seeks to explain some aspect of our psychology — and an approach that is becoming increasingly prominent — focuses on character traits. There is a philosophical tradition, traceable back to Aristotle, which treats morality not as a matter of following rules or of consequences of actions but rather of acquiring and exercising certain character traits, the virtues. Aristotle’s list of virtues included courage, truthfulness and generosity. These traits are virtues, according to Aristotle, because to acquire and exercise them is to fulfil one’s function as a human being. Aristotle’s view of functions was teleological, relying on the idea of purpose in nature. Rejecting the teleology inherent in Aristotle’s moral philosophy, many contemporary moral philosophers have

---

56 Thanks to Hatha McDivitt for the memorable name.
57 It is wearying to walk over such well-trodden ground. That an evolutionary explanation for why we are nice and helpful does not amount to an explanation for morality was appreciated at least as far back as 1952, by C.S. Lewis of all people, in Mere Christianity:

[Some people wrote to me saying, “Isn’t what you call the Moral Law simply our herd instinct and hasn’t it been developed just like all our other instincts?” Now I do not deny that we may have a herd instinct: but that is not what I mean by the Moral Law. We all know what it feels like to be prompted by instinct - by mother love, or sexual instinct, or the instinct for food. It means that you feel a strong want or desire to act in a certain way. And, of course, we sometimes do feel just that sort of desire to help another person: and no doubt that desire is due to the herd instinct. But feeling a desire to help is quite different from feeling that you ought to help whether you want to or not (2001/1952: 9).

Lewis thought that it was necessary to appeal to the existence of God in order to explain our sense that some things are good, right, obligatory, etc. and others bad, wrong, forbidden, etc. While I don’t think his argument for the existence of God works, I am certainly sympathetic to the way in which he sets it up!
58 I am relying here on Irwin’s (1999) translation of the Nichomachean Ethics.
continued to think of character traits as the locus of morality. Some recent work on the evolution of morality has attempted to update Aristotle’s virtue ethics.

Arnhart (1998) and Casebeer (2003), whose work was mentioned in Chapter 1 (see §1.1.2e), offer this kind of evolutionary explanation of morality. Jonathan Haidt also offers an evolutionary explanation of morality that is intended to connect up with virtue ethics (Haidt & Joseph [2004]; [2007]). Most importantly for my purposes, Miller’s (2007) development of SSM proceeds in terms of virtue ethics and the evolution of character traits. In what follows, I will devote some time to presenting and critiquing Miller’s development of SSM in these terms. My aim in doing so is to set up a fix for SSM. As mentioned above, the fix will be to combine SSM with a kind of evolutionary explanation of morality that does get the explanatory target right. The details of that kind of approach will follow my critical discussion of Miller’s updated version of SSM.

Chapter 1 raised the worry that Miller’s initial version of SSM qua evolutionary explanation of morality did not get the explanatory target right (see §1.1.2a). Miller’s updated version of SSM is framed in terms of moral virtues and virtue ethics. “Many human virtues,” Miller claims, “evolved through sexual selection as costly signals” (2007: 100). Moral virtue, according to Miller “emerges from the interaction of traits and preferences” (2007: 103). By way of example, he says “kindness is a moral virtue because it predicts specific prosocial behaviours, and is valued as such” (2007: 108). Miller’s reason for framing SSM in terms of virtues is that doing so “provides a useful counterbalance to the traditional consequentialist (utilitarian, payoff-based) ethics that have influenced previous evolutionary theories of altruism (e.g. kin selection and reciprocal altruism)” and also “shifts the level of analysis usefully from isolated altruistic acts to stable personality traits” (2007: 99).59

59 Miller’s use of ‘altruism’ and ‘altruistic’ is rather haphazard. Sometimes he means helpfulness, sometimes fitness sacrificing, and sometimes altruism as I have defined it above. I will attempt to disambiguate Miller’s usage as I go along here.
There are, it seems, two things that make up a moral virtue on Miller’s view: a stable personality trait and a judgement, preference, or evaluation (he uses all three terms at various points) regarding that trait. This updated version of SSM is an advance over Miller’s initial version, at least insofar as virtues thus understood constitute a unified explanatory target, unlike the miscellany of behaviours discussed in The Mating Mind. Furthermore, stable personality traits seem like the right kind of thing for which to seek an evolutionary explanation, unlike individually or culturally idiosyncratic behaviours. Even so, this way of framing SSM remains problematic.

Miller’s justifications for framing SSM in terms of virtue ethics are shaky. Firstly, it is not obvious how consequentialist ethical theories have influenced discussions of the evolution of morality. Consequentialist moral theories and the theories of kin selection and reciprocal altruism do both focus on payoffs but the currency in circulation in the two cases is quite different. In the latter case, it is fitness, while in the former, it is whatever is supposed to be of value. The resemblance here is superficial, the connection tenuous at best. Secondly, Miller’s reason for thinking moral evaluation is best done at the level of persons rather than at the level of acts is that we evaluate acts in context, but this is nothing that a proponent of act-level evaluation need deny. What Miller says is sensible enough but does not suffice to motivate a shift away from judging acts to a different level – the “moral-person level” (2007: 103) – of evaluation.

Perhaps there is a more substantial point to be made in the vicinity of Miller’s first justification. We might note that there is no project of giving an evolutionary explanation for why there are utilitarian moral agents, since there are no such agents. Focusing on virtue ethics rather than utilitarianism (or some other form of consequentialism) when giving an evolutionary explanation of morality is therefore preferable because there really are agents

---

60 There is no reason that such evaluation cannot take into account the capabilities of the agent, the features of the situation, and the relationships between the agent and those affected by the act.
61 Not even the arch-utilitarian Peter Singer. See Bailey (2000) for an interview with Singer in which the philosopher’s moderately luxurious lifestyle and failure to euthanize his ill mother are (rather too gleefully) emphasized.
who exhibit and prefer virtues (as Miller understands them). This is a point worth making: virtues, unlike utilitarians, constitute a genuine explanatory target. As with the various approaches to the evolution of morality discussed above, though, we can ask whether the target being aimed at is really morality. I turn now to that question.

Recall, on Miller's view, that a moral virtue is a stable personality trait plus a preference, judgement, or evaluation regarding that trait. Miller stresses that he:

...do[es] not assume that the "virtues" historically identified by philosophers will equal the moral adaptations that can be identified in humans...[or] that the idealistic reasons for advocating certain virtues in normative ethics will have anything to do with the selection pressures that may have shaped those virtues phylogenetically (2007: 99).

It is legitimate to ask, then, why the character traits Miller thinks evolved via sexual selection should be considered moral virtues. Of course, some of the traits Miller claims have evolved by sexual selection – for example kindness, fairness and honesty – are widely accepted as such. However, Miller also counts as moral virtues community solidarity, cultural pride, and (for men) manliness and (for women) femininity. As he acknowledges, some think those traits are properly considered vices instead of virtues (the vices of authoritarianism, jingoism, and sexism). Furthermore, Miller includes intelligence, humour, health and beauty among the moral virtues produced by sexual selection. Notice, the last two items on that list are not even character traits, let alone things that are typically imbued with moral significance.

Miller anticipates the scepticism likely to greet some of his claims about what counts as a moral virtue. He has this to say in his defence:

If we are seeking a descriptive explanation for human morality, we should attend to the person-perception judgements that may have causally driven moral evolution in our species. Ultimately, it is an empirical question whether ordinary people judge traits to have a moral or quasi-moral status (2007: 109).
Miller’s focus on judgements here is an important step forward. Preferences for certain character traits, such as kindness and courage, might help explain the evolution of those traits. That may be an important project in its own right (cf. §2.2.3), but absurdity lurks just beyond the supposition that the existence of preferences for certain traits qualifies those traits as moral virtues. This would count height, wealth and youth as moral virtues, which is clearly wrong even if there are widespread preferences for such things in a potential mate. If judgements are to be partly what defines virtues, though, not just any kind of judgement will do. For instance, judging that heroism is sexy won’t do to substantiate the claim that heroism is a moral virtue. The judgements in question had better be moral judgements. This may seem obvious but it is obscured in Miller’s discussion by his talk of what we prefer or find attractive.

Now we come to the crunch. Repeatedly in my discussion of various approaches to explaining the evolution of morality, a distinction has been drawn between things we judge morally and the very capacity to make moral judgements. Some researchers have focused on explaining the evolution of certain behaviours we tend to judge in a positive moral light, such as helping. Other researchers have focused on explaining how evolution could instil in us what we often take to be morally laudable motives, prominently including altruistic ones. Yet others, Miller among them, seek to explain the evolution of character traits widely considered morally worthy. All these researchers focus on explaining some target of our moral judgements, rather than on the trait of making moral judgements itself. In the next section, I will lay out what I take to be the right kind of approach to explaining the evolution of morality, namely, the kind of approach that seeks to explain the evolution of that trait.

2.2.3b. Revising SSM: Miller plus Joyce

Some purported evolutionary explanations of morality are explanations for our tendency to make moral judgements. The work of several researchers falls into this class. I have already mentioned (in §1.2.2e) Ruse & Wilson’s claim that “our sense of ‘right’ and the corresponding sense of ‘wrong’ [are] brought
about by ultimately biological processes" (1986: 179). Ruse developed this claim in numerous subsequent pieces and the quote below nicely exemplifies the kind of evolutionary explanation of morality that focuses on our tendency to make moral judgements:

The key hypothesis is that evolution has made us innately inclined to think in certain ways... We think that we ought to help, that we have obligations to others, because it is in our biological interests to have these thoughts...We are moral because our genes, as fashioned by natural selection, fill us full of thoughts about being moral (Ruse [1993: 503-504]).

This kind of view has also been advanced by psychologists Leda Cosmides & John Tooby, whose work aims to show how “evolution [has] equipped our minds with moral heuristics: decision rules that generate intuitions about fairness and justice, punitiveness and approval, right and wrong” (2006: 175). Marc Hauser (2006) is yet another who pushes this line, claiming that we have an evolved “moral organ,” a dedicated set of neural circuits designed to deliver verdicts of right and wrong.

Cosmides & Tooby and Hauser commit themselves to very specific accounts of our capacity for moral judgement. For Cosmides & Tooby, that capacity is the result of our possessing a number of specialised mental mechanisms for various kinds of moral thought, such as the so-called “cheater-detection” module for reasoning about violations of rules governing social exchange (Cosmides & Tooby [2004: 91]). For Hauser, our capacity for moral judgement is much like our capacity for language as described by Noam Chomsky: we are born with a “universal moral grammar,” a collection of largely unconscious, abstract (i.e. lacking in specific content) principles whose parameters are fixed by our early developmental environment.

A commitment to the massive modularity of mind thesis or to a principles-and-parameters picture of morality, or indeed to any particular view of how our capacity for moral judgement works, is not necessary for adopting this kind of approach to explaining the evolution of morality. Richard Joyce, for example, offers an evolutionary explanation of morality that targets “the trait of making moral judgements” (2006: 17) but remains relatively
uncommitted on how that trait is realised in us. Joyce’s claim is an ultimate one about the adaptive value to our ancestors of making moral judgements. Proximate-level claims about how the capacity to make moral judgements is implemented in us do feature in his discussion but are not its focus.62

Since I am proposing to combine SSM with Joyce’s account, I will lay out the key features (for my purposes) of Joyce’s account. Specifically, I will lay out his views on the nature of moral judgements and on why a tendency to make such judgements would have been favoured by selection.

Moral judgements, Joyce claims, are considered by everyday, ordinary folk moralizers to be inescapable and authoritative. Take the moral judgements ‘doing \(A\) is obligatory’ and ‘doing \(B\) is forbidden.’ Joyce claims that such judgements are taken to apply (when they do) regardless of individuals’ wants or needs. The former applies to Bob even if he really doesn’t want to do \(A\) and the latter applies to Jill even if she really needs to do \(B\). In this respect, moral judgements are unlike judgements about what is prudent, which apply only in virtue of individuals having particular wants and needs. In this respect, moral judgements resemble judgements like ‘doing \(X\) is polite’ and ‘doing \(Y\) is pious.’ What marks moral judgements off from judgements about politeness and piety, Joyce says, is that moral judgements are taken to give everyone a reason to comply with them. Were Jim to guzzle communion wine at an Alcoholics Anonymous meeting it would be impious and rude to boot. If Jim happens to be an atheistic social iconoclast, these judgements would not be retracted, but it would be allowed that they give Jim no reason not to swill away. By contrast, Joyce says, folk moralizers do not take there to be any way of legitimately ignoring moral judgements.63 Joyce dubs the conjunction of inescapability and authority “practical clout” (2006:

---


63 It is important to keep questions about (a) the nature of moral judgments vis a vis their supposed inescapability and authority distinct from (b) questions about whether the inescapability and authority of moral judgments can be substantiated or justified. My focus at this point is on (a): it is the character of moral thinking that is relevant to Joyce’s claim (and Ruse’s for that matter) about the evolution of the capacity for such thinking.
58). He considers the fact that folk moralizers imbue moral judgments with practical clout to be an important one, and one for which an evolutionary explanation of morality must account.

A second important feature of moral judgements is, Joyce says, their content: “human morality takes interpersonal relations as its central subject matter” (2006: 66). Joyce circumspectly refuses to rule out the possibility of moral systems that have nothing to do with interpersonal relations. He also allows that some moral judgements are purely self-regarding. Still, he insists, insofar as one is interested in explaining the evolution of human morality, it is interpersonal relations, especially considerations of harm and fairness, which take centre stage. Joyce also observes that “the actions that morality prescribes with categorical force are those that constitute or promote, roughly speaking, cooperation [i.e. helpfulness]” (2000: 714).

As was stressed above, helpfulness and morality cannot be equated. Not all helpful behaviour is moral, or all moral behaviour helpful. Still, there is undeniably considerable overlap between helpfulness and what we consider morally required or at least laudable behaviour. This observation is the beginning of Joyce’s explanation of why the trait of making moral judgements would have been fitness-enhancing for our distant ancestors. As will be discussed in detail in my next chapter, there are numerous reasons to think that helpfulness would have been fitness-enhancing for our distant ancestors. Joyce emphasises the various evolutionary pressures favouring helpful behaviour, and then claims that the tendency to make moral judgements would have been favoured by selection because moral judgements, in virtue of their “practical clout,” help motivate fitness-enhancing helpfulness.

---

64 Without going into detail at this point, it is worth stressing that the helpfulness in question is not indiscriminate. The claim is that under certain circumstances and to some degree, helping others was fitness-enhancing for our distant ancestors.

65 Ruse says (1998: 222) that “because it is biologically advantageous for us to help and cooperate, morality...has evolved to guide and stiffen our will.” I take this idea to be central to the kind of evolutionary explanation of morality I am considering in this section. I take Joyce’s work to be a particularly careful and detailed cashing out of this idea.
Joyce does not deny that helpful behaviour can be motivated by other-regarding emotions such as sympathy or even by cold calculation of self-interest. His claim is only that moral judgements bolster motivations to be helpful, giving agents with the tendency to make them something of an edge over agents equipped only with emotions and prudential concern:

[A]n individual sincerely judging some action in a morally positive light increases the probability that the individual will perform that action (likewise, mutatis mutandis, judging an action in a morally negative light) (2006: 109).

In essence, then, Joyce claims that selection for certain kinds of helpful behaviour could have also favoured the tendency to make moral judgements, since making such judgements makes acting in those helpful ways more likely. It seems open to an advocate of SSM to make a parallel claim. SSM could be stated as that claim that inter-sexual selection in favour of certain kinds of signalling behaviours could have also favoured the tendency to make moral judgements, since making such judgements makes acting in those attractive ways more likely.

Framing SSM in the manner just suggested would not be entirely alien to Miller’s way of thinking. In his initial presentation of SSM, Miller posed the following (largely rhetorical) question: “Why do we consider such displays [i.e. displays of generosity, fairness, kindness, heroism] particularly “moral,” as compared with other courtship displays?” (2000: 308). Why is it, Miller wonders, that we judge some sorts of behaviour to be not just attractive but required, not just off-putting but forbidden? Emphasising the motivational efficacy of judging in moral terms helps provide an answer to this question.

Admittedly, framing SSM as a hypothesis about the evolution of the tendency to make moral judgements in the way I am suggesting would add another empirical burden to the load that SSM must bear. SSM thus framed depends on the truth of the claim that judging in moral terms really is motivationally efficacious over and above experiencing emotions and calculating self-interest. The empirical cost of framing SSM in the way I am
suggesting is, however, worth incurring. Doing so picks the hypothesis out from among the many that fall foul of what I have called the synecdoche fallacy. It turns SSM into a claim that, if true, would realise the possibility mentioned in Chapter 1 (see §1.1) that evolutionary theory might help reveal “how people come to acquire ethical concepts, to make ethical judgements about themselves and others, and to formulate systems of ethical principles” (Kitcher [1985: 417]).

A fair question to ask at this point would be why focussing on the tendency to make moral judgements does not itself commit the synecdoche fallacy. After all, it remains the case that behaviours, rules, character traits, motives and the like are all important elements of human morality, considered as a natural phenomenon to be explained in some way or other. The answer, I think, is that this approach holds out the promise of explaining why we think in moral terms at all. This seems a particularly fundamental question. The approaches considered above focus on some or other element of the phenomenon that is human morality. This approach tries to say why there exists such a phenomenon in the first place.

In the next section, I will defend Joyce’s approach to explaining the evolution of morality. Before moving on, though, I should flag a pair of issues with the combination of Joyce’s view and SSM that I have proposed. The first is a potential confusion that needs clarifying. The second concerns the distinctiveness of SSM as a hypothesis about the evolution of morality, framed in the way I have proposed.

The phrase ‘moral judgement’ can be used to refer to a public utterance or to a private mental event. It is moral judgements in the second sense that play the important motivating role in Joyce’s account of the evolution of morality. Miller, however, suggests that moral judgements are themselves among the costly signals produced by sexual selection (2000: 340). Miller must be using ‘moral judgement’ in the first sense here, since it is only in that sense that moral judgements could possibly play a signalling role. It is important not to confuse these two claims about the evolutionary function of
making moral judgements. The two claims are not incompatible, of course. I will discuss in detail in Chapter 4 the plausibility of Miller’s claim that moral judgements *qua* public utterances function as costly signals in the context of mate choice.

Miller might perhaps worry that SSM as stated above is in danger of shrinking to a footnote on Joyce’s claims about the evolution of morality. To this worry, I would say firstly that proposing (and providing evidence for) another kind of benefit (i.e. a mating benefit) to motivationally resolute moral thinking is not a trivial contribution to explaining the evolution of morality. I think SSM is an important part of a cumulative case for the selective advantage of that trait. Also, Miller’s claim that SSM is a novel, distinctive kind of explanation for the evolution of morality depends on the claim that sexual selection is a somehow special sort of evolutionary process. Whether that latter claim can be substantiated, and what becomes of SSM’s claim to novelty if it cannot, will be discussed in Chapter 3.

### 2.3. Defending Revised SSM

In this section, I will discuss a pair of criticisms of Joyce’s approach to explaining the evolution of morality. A word about my strategy in this section is in order. The criticisms I consider come from Stephen Stich and are directed at Joyce specifically. However, Stich’s criticisms draw on a wide range of work by others in philosophy and empirical disciplines including psychology and anthropology; they are not worries peculiar to him. Furthermore, the criticisms apply not just to Joyce but to the general approach Joyce takes, which is also taken by Ruse among others. Addressing these criticisms is of interest beyond just defending Joyce’s work. Importantly for my purposes, some reply to Stich’s criticisms is needed in defence of the modified version of SSM I have proposed.

In what follows, I will consider Stich’s claims that moral judgements as understood by Joyce (and as I am understanding them here) are not human universals and that a focus on helpfulness in explaining the evolution of
morality is overly narrow. Neither criticism can be easily disposed of, so while there are replies to be made and the upshot is (I think) a happy one for Joyce and my modified SSM, getting there will take some doing.

2.3.1. ‘Joyce-style’ Moral Judgements

It is central to Joyce’s account of the evolution of morality that moral judgements have clout, that the demands of morality are considered inescapable and authoritative. It is in virtue of their clout that moral judgements are motivationally efficacious in a way that could drive selection for the tendency to make such judgements. In response to Joyce’s work, Stephen Stich has questioned whether the making of moral judgements as Joyce conceives of them is really a human universal (the kind of trait for which an evolutionary explanation could sensibly be sought). Stich voices his suspicion that:

…the practice of making moral judgements of the sort that Joyce describes is a culturally and temporally local one restricted to Western (and Western-influenced) cultural groups in relatively recent times (2008: 234).

One of Joyce’s main lines of evidence for the claim that making moral judgements really is a human universal is the research in moral psychology on the moral/conventional distinction (Smetana [1981]; Turiel [1983]). Stich’s denial that the tendency to make ‘Joyce-style’ moral judgements is a human universal relies heavily on empirical work by him and his collaborators (Kelly et al. [2007]) challenging the moral/conventional distinction. Thus, it is worth taking a moment to consider research in the moral/conventional task tradition in detail.

Elliot Turiel proposed that moral rules and conventional rules are distinguished (in practice, by ordinary folk) along several dimensions: scope, dependence on authority, seriousness of violation, and grounds or justification. According to the Turiel tradition, moral rules are thought to have universal scope and to hold independently of the pronouncements of authorities. That is, moral rules are seen as holding in all places and times rather than just locally,
and breaking moral rules is still seen as wrong even if doing so is explicitly permitted by an authority. Furthermore, violating moral rules is considered relatively serious, compared to violating conventional rules (which will be detailed in a moment). Finally, Turiel claimed, justifications for moral rules typically appeal to the harms or injustice suffered by victims when those rules are violated. Turiel proposed that conventional rules, by contrast, are treated as having limited scope (holding only locally) and as authority-dependent (such that breaking the rule is not wrong if permitted by an authority). Violations of conventional rules are seen as less serious than moral violations and, lastly, justifications for conventional rules tend to cite ‘minimising disruption’ or ‘maintaining social order’ as relevant considerations rather than harm or injustice.

The moral/conventional task has been administered to a wide range of subjects. Turiel’s proposals have apparently received striking support. Subjects presented with paradigmatic cases of moral wrongdoing (e.g. one child gratuitously pulling another’s hair in the classroom) judge such things to be wrong and maintain that judgement even when asked to imagine that the act occurred in a different setting (e.g. a classroom in a different country) or that an authority figure (e.g. a schoolteacher) explicitly permitted the act. By contrast, when asked to judge paradigmatic cases in which conventions are violated (e.g. a child wearing pyjamas instead of a uniform to school), subjects allow that such acts might not be wrong everywhere and that such acts would not be wrong if they were permitted by an authority. The weight of evidence seems to indicate that rule violations involving harm or injustice evoke a “signature moral response pattern,” in which “rules are judged to be authority independent and general in scope; violations are more serious, and rules are justified by appeal to harm, justice or rights” (Kelly et al. [2007: 119-120]).

66 Subjects have varied in age, cultural background and religious affiliation, as well as mental capacity (including, for example, psychopaths and sufferers of autism). See Kelly & Stich (2008: 354) for references.

67 Turiel’s proposals regarding differences in the perceived serious of moral versus conventional violations and in the typical justifications offered for the two kinds of rules have also been supported by this large body of empirical work.
Joyce notes that work in the moral/conventional task tradition, especially concerning the supposed authority-independence of moral rules, "meshes well" with his claims about the practical clout of moral judgements (2006: 136). Joyce is also at pains to point out that the idea of moral judgements having putative inescapability and authority is not some philosopher's fancy but rather a feature of common thought and talk about morality ([2006: 61, 193]; see also [2008: 246-247]).

It is thus important to Joyce's case that the moral/conventional distinction hold up under the assault by Stich and his colleagues.

Kelly et al. (2007) deny that all transgressions involving harm or fairness evoke the signature moral response pattern described above. Their case against this claim begins with the observation that researchers working in the moral/conventional task tradition have presented subjects with a very limited range of stimuli. Moral/conventional task studies have typically asked subjects to judge schoolyard variety harms like hair pulling. Kelly et al. created more complex harm scenarios in order to test whether subjects' responses to a wider range of stimuli would accord with previous findings.

I will give a couple of representative examples in some detail here, since the specifics matter when evaluating the results and interpretation of the study. Kelly et al. constructed contrasting pairs of harm scenarios that varied with respect to either where the harm described occurred, when the harm described occurred, or whether the harm described was sanctioned by an authority. One of the scenario pairs used by Kelly et al. was "Whipping/Time" (2007: 123-4). This scenario pair was meant to test whether subjects' moral

---

68 I am not equating or suggesting that Joyce equates inescapability with universality of scope or authority with authority-independence. The results of moral/conventional task research are friendly to Joyce's claim about moral judgments having practical clout but they do not establish it. Conversely, undermining such research would not undermine Joyce's claim. Whether studies that specifically investigated inescapability and authority would turn up results consonant with Joyce's view is an interesting question that will hopefully be addressed by future research.

69 The attack on the moral/conventional distinction in Kelly et al. (2007) has been taken seriously by several prominent philosophers and psychologists (e.g. Haidt & Joseph [2007: 372]; Prinz [2007b: 282]; Doris & Nichols [forthcoming]), meaning the issues I discuss above are likely to be of interest outside the context of Stich's criticism of Joyce.
judgements really were temporally universal in scope. Subjects who received this scenario first saw a version that ran as follows:

Three hundred years ago, whipping was a common practice in most navies and on cargo ships. There were no laws against it, and almost everyone thought that whipping was an appropriate way to discipline sailors who disobeyed orders or were drunk on duty.

Mr. Williams was an officer on a cargo ship 300 years ago. One night, while at sea, he found a sailor drunk at a time when the sailor should have been on watch. After the sailor sobered up, Williams punished the sailor by giving him 5 lashes with a whip.

Subjects were then asked to respond to a different version:

Mr. Adams is an officer on a large modern American cargo ship in 2004. One night, while at sea, he finds a sailor drunk at a time when the sailor should have been monitoring the radar screen. After the sailor sobers up, Adams punishes the sailor by giving him 5 lashes with a whip.

Another scenario pair used by Kelly et al. was “Military/Authority” (2007: 125). This scenario pair was meant to test whether subjects’ moral judgements really were authority independent. Subjects given this scenario were first shown the following material:

For many years, the military training of elite American commandos included a simulated interrogation by enemy forces in which the trainees were threatened and physically abused. Most people in the military believe that these simulated interrogations were helpful in preparing trainees for situations they might face later in their military careers. Though no one was ever killed or permanently disabled by the physical abuse they received during these simulated interrogations, the trainees often ended up with bruises or injuries that lasted for a week or more.

Recently, the Pentagon issued orders prohibiting physical abuse in military training. Sergeant Anderson is a soldier who trains elite American commandos. He knows about the orders prohibiting physical abuse and his immediate superiors have ordered him not to do it. Nonetheless, he regularly threatens and physically abuses trainees during the simulated interrogations that he conducts.

Subjects were then shown this:

Now suppose that the Pentagon had never issued orders prohibiting physical abuse in military training, and that Sergeant Anderson’s superiors had told him that the use of physical abuse was acceptable in simulated interrogations.
Subjects were asked to indicate whether causing the described harm was OK or not (a yes/no question) and also to rate the behaviour of the agent who caused the harm (on a 9-point scale running from ‘not at all bad’ to ‘very bad’). Kelly et al. report a significant difference in the number of subjects who judged whipping to be OK in the past and the number of subjects who judged whipping to be OK now: “many subjects think whipping was OK 300 years ago but they do not think it is OK now” (2007: 126). That is, Kelly et al. say, a significant number of subjects made moral judgements that were temporally local, judgments that were not generalised across different time periods (2007: 129). Also, while more than half of subjects responded that physical abuse in military training was OK when not prohibited, very few judged it to be OK when prohibited, showing, Kelly et al. claim, that “many subjects do not judge the harmful transgression to be authority independent” (2007: 127).

Kelly et al. claim their study casts doubt on the idea that there is a signature moral response pattern evoked by rule violations involving harm. In particular, moral judgements about harms of other than the simple schoolyard variety need not exhibit authority independence or universality of scope. Notice, while the harm scenarios Kelly et al. described were more complex than schoolyard variety ones, they were not entirely outlandish or even particularly far-fetched. If ordinary folks’ moral thinking about such cases does not exhibit authority independence and universality of scope, then Joyce’s claim that common thought and talk about morality supposes moral judgements to have ‘clout’ may well be imperilled.

Kelly et al. end their paper by asking:

Is there something special about these simple harm transgressions [i.e. the ‘schoolyard’ harms described in previous studies] that is not shared by the more ‘grown-up’ transgressions that we also used in our study? (2007: 129)

70 A structurally similar scenario pair produced similar results, Kelly et al. report, for geographical differences, with subjects making judgments that were not generalized to cover harms occurring in other places.
The answer, I think, is that there is a relevant difference between the schoolyard harm scenarios used in previous moral/conventional task studies and the harm scenarios Kelly et al. use.\textsuperscript{71} The former describe harms that are very clearly transgressions, that is, wrongful harms, ones that ought not be inflicted on the victim. In the very simplicity of schoolyard harms lies their strength as experimental probes. Early research on the moral/conventional distinction used examples like pulling hair and pushing people off swings precisely because those are paradigm cases of harmful moral violations. Subsequent moral/conventional task studies continued to employ these simple cases for that reason.\textsuperscript{72}

At risk of pointing out the obvious, I think it worth stating that not all harms are considered transgressions. Some harms are considered morally justified, such as the harm inflicted by a doctor or dentist in treating a patient. Some harms may even be seen as morally required, such as the harms involved in meting out deserved punishment to wrongdoers. In general, whether a given harm counts as a transgression will depend in part on the context in which the harm is inflicted. It thus seems sensible to ask whether, across the contrasting scenario pairs Kelly et al. constructed, it is the temporal differences (or the differences in authoritarian approval) \textit{per se} that make for the differences in subjects' responses.

\textsuperscript{71} In what follows, I offer a largely speculative response to Kelly et al. For a recent criticism of Kelly et al. that focuses on methodological issues to do with data analysis, see Paulo Sousa (2009). Sousa argues that the way in which Kelly et al. analysed their data made their findings seem more striking than they really were. Kelly et al. compared the total number of 'yes' answers to the 'Is it OK...?' question for each scenario with the total number of 'yes' answers to the 'Is it OK...?' question for the contrasting scenario. For example, in the 'Military/Authority' scenario pair, Kelly et al. report that only 9% of subjects said physical abuse in military training was OK when it was prohibited but the percentage of subjects saying physical abuse in military training was OK 'jumped' to 58% when physical abuse in military training was permitted by military authorities. Sousa points out (rightly I think) that, if the question is whether subjects are making authority-dependent or authority-independent moral judgments, then what really matters is how many subjects said that physical abuse in military training was not OK when prohibited by authority \textit{and} said that physical abuse in military training was OK if it was permitted by authority. In Sousa's re-analysis of the data, subjects like \textit{that} are in the distinct minority for many of the scenario pairs. For some scenario pairs, though, such subjects are common. My discussion above can be read as a way of accounting for these subjects' responses without having to suppose their moral judgments are authority-dependent (or limited-scope, in the case of the other scenario pairs).

\textsuperscript{72} As well as to facilitate comparison of results across subject populations, e.g. children and psychopaths.
Recall, a significant proportion of subjects in the Kelly et al. study judged that it was OK to whip derelict sailors 300 years ago but not OK to whip derelict sailors now and that physical abuse in military training is OK when not prohibited but not OK when it is prohibited. Recall also, Kelly et al. claim that what is going on here is that subjects are making moral judgements about harm that are (temporally) limited in scope and authority-dependent. However, the very complexity of the harm scenarios used may be part of the explanation for the pattern of results Kelly et al. found.

Suppose subjects make their particular moral judgements, such as ‘whipping Mr Williams in 2004 is not OK’ and ‘whipping Mr Williams 300 years ago was OK,’ by applying to the particular cases described a more general moral judgement to the effect that (roughly) ‘inflicting harm is permissible when necessary as means to achieving a compensatingly valuable end.’ That such a general moral judgement might be in play is quite plausible. After all, something very like it no doubt accounts for the observations above about common views on when dentists and executioners can permissibly inflict harm on others.

The general moral judgement about when harm is OK that I just sketched takes circumstances into account. Perhaps subjects reading about Mr Williams being whipped 300 years ago think that in those wild times on the high seas, whipping may have been the only punishment fierce enough to deter sailors from drinking on duty and thus the only way to safeguard the lives that could depend on a watchful lookout’s warning. In today’s setting, by contrast, whipping is probably not seen as necessary to maintaining shipboard discipline. If this speculation is on track, then the general moral principle sketched above would generate different particular judgements when applied to the contrasting scenarios described in the study by Kelly et al.

I am tempted by a similar speculative account of what is going on when subjects judge that physical abuse in military training is OK when not prohibited but not OK when it is prohibited. Suppose again that subjects make their particular moral judgements by applying something like the general
principle sketched above. Suppose in addition that subjects take the pronouncements of relevant authorities to be a good guide to the necessity or otherwise of certain harms in bringing about valuable ends. This further supposition is by no means unreasonable. Deference to (supposed) experts and epistemic division of labour is a feature of our everyday practices, in moral judgement no less so than in other arenas. If these two suppositions are correct, it would account for the responses of subjects who judged that physical abuse in military training is OK when permitted by military authorities but not when it is prohibited.

Notice that on my speculative account, it is not necessary to suppose that subjects are making limited-scope, authority-dependent moral judgements. It is not the differences in temporal location or authoritarian approval that make the difference to subjects' responses, on my account. Those things are relevant, to be sure, but relevant to the application of a general moral judgement that we have no reason to doubt is considered universal in scope and independent of authority. (Whether inflicting a given harm is permissible, subjects think, depends on whether the harm is a necessary as means to achieving a compensatingly valuable end, and this is true regardless of where or when the harm occurs or whether some authority happens to sanction or prohibit it.)

I have made no bones about the speculative nature of my discussion here. Nor will I now pretend to have settled anything. Stich's objection to Joyce's claim that moral judgements are characterised by putative inescapability and authority was motivated in part by empirical work he took to seriously challenge to moral/conventional distinction. I have offered some reasons to doubt that we must interpret that work as Stich does. Mine is one small move in a complicated debate.

---

73 It would be remiss of me not to mention that what Sousa calls the "Law Scope Problem" (2009: 227-229) touches on some of the issues with the Kelly et al. study that I discuss above. We are, I think, getting at the same problem from different angles.
In the next section, I will move on to another criticism of Joyce’s approach to explaining the evolution of morality (one also pressed by Stich, among others). As in this section, my aim will be to defend the elements of Joyce’s view that are particularly relevant to the modified version of SSM I have proposed in this chapter.

2.3.2. The Extent of the Moral Domain

As well as questioning whether moral judgements have the particular normative form Joyce supposes, Stich questions Joyce’s idea that “the primary sort of behaviour moral judgement was selected to motivate is cooperative or prosocial behaviour” (2008: 234). He notes the prevalence across different cultures and time periods of norms that seemingly have little to do with helping, harming or fairness, such as norms about what clothing can be worn, what foods can be eaten, how to show deference to high ranking people, and how to dispose of the dead. He then says:

Since norms governing all of these matters are as ubiquitous as norms governing reciprocity, it strikes me as rather implausible that reciprocity and prosocial norms should have pride of place in an account of the evolution of morality (2008: 235).

This claim strikes me as deeply puzzling. According reciprocity pride of place in an explanation of morality despite the existence of many other kinds of norms would strike one as implausible (surely?) only if one took those other norms to also be moral norms. But Stich is very well aware that the issue of which norms are moral norms is a vexed one. Indeed, his discussions of this issue are nuanced and useful (Nado et al. [2009: 619-21]; Sripada & Stich [2007]). Stich does not assume that norms about clothing, food, corpses and so on are moral norms; he is officially agnostic on just how the category of moral norms is to be defined (and even whether such a category can be defined at all). So, since Joyce’s claim is not that reciprocity has pride of place in an explanation of norms simpliciter, it is not clear to what Stich is objecting here.

74 Just how the issues Stich raises confront a Joyce-style approach to the evolution of morality will be considered below, after the issues are put on the table.
What reason is there to think that morality is about more than interpersonal relations, including especially issues of harm and fairness? Perhaps a good place to begin this discussion is with an examination of the empirical work meant to support the claim. A key figure here is Jonathan Haidt (see e.g. Haidt et al. [1993]; Haidt [2001]; Haidt & Joseph [2004]; [2007]; Haidt [2007]). Haidt claims that there are four moral “domains,” meaning (so far as I can tell) broad classes of actions that are treated as morally evaluable. These are the domains of suffering (concerned with harm), reciprocity (concerned with fairness), hierarchy (concerned with respect and loyalty), and purity (concerned with pollution). Only the first three of these four domains are concerned with interpersonal relations, which, as Haidt notes, may strike one as strange:

There is an odd corner of moral life, odd at least for modern Westerners, who tend to think of morality as strictly concerned with how we treat other people. That corner is the profound moralisation of the body and bodily activities, such as menstruation, eating, bathing, sex, and the handling of corpses (Haidt & Joseph [2004: 60]).

Haidt claims that there are cultures in which morality is not focussed centrally on issues of harm or fairness, being based instead around the ideas of purity and pollution. A key piece of empirical work supporting this claim is Haidt et al. (1993), in which subjects of different cultures and socioeconomic standing (SES) were studied in order to determine whether culture and/or SES influence the kinds of issues people treat as moral issues (the “domain of morality,” as Haidt et al. put it). In particular, the study investigated people’s responses to harmless but disgusting or disrespectful actions. To ensure that judgments about harmfulness were not confounding his results, Haidt asked subjects whether they thought the act they were being asked to judge was harmful (to anyone, including the agent) or not. Only data from subjects who judged the acts in question to be harmless (which was the great bulk of them) was considered. In this respect, Haidt’s work is superior to that of Shaun

---

75 Figuring out what psychologists, anthropologists, and other non-philosophers mean is often a large part of the challenge in attempting to do empirically-informed philosophy. Unfortunately, philosophers’ glosses on the empirical literature sometimes amount to little more than a change of jargon. According to Owen Flanagan (2007: 127), Haidt’s moral domains are “domains of life...that humans build normative space around.”
Nichols (2002), who surveyed subjects’ responses to moral, conventional, and disgusting transgressions but did not include such a question in his study design, leaving open the possibility that harmfulness played an unrecognised role in generating the pattern of results he found.

In considering Haidt’s study, it is crucial to be clear about how Haidt et al. operationalised what it is to treat something as a moral issue. Haidt et al. contrasted two “stances” that might be adopted toward an act: a “moralising” stance and a “permissive” stance. To adopt a moralising stance toward an act is to “endorse interference” and to “universalise” one’s judgement that the act is wrong, that is, to judge that people should be prevented from performing the act and punished if they do so and, furthermore, to judge that performing the act is wrong everywhere, not just locally (1993: 613). To adopt a permissive stance is to fail to make one or both of those judgements about the act.

Notice that, on the account just laid out, one would count as taking a permissive stance toward an act if one judged that it is wrong and should be prevented/punished but only locally. Suppose a headhunter believes that his fellow tribesmen must take heads after battle and should be punished (perhaps by beheading) for failing to do so. Suppose he also believes that the foothunters from across the valley are under no such obligation and do not deserve punishment for failing to take heads after battle. This fellow would count as permissive with respect to taking heads. This would probably come as a shock to the folk he has beheaded for failing to take heads, were those folk in a position to be shocked by anything. Calling this headhunter ‘permissive’ is terminologically jarring at least. There may also be a deeper problem, though. Haidt may be ruling out the very possibility of moral relativism being true. If one thinks the wrongness of acts is always a local and culturally-specific phenomenon, one cannot count as moralising by Haidt’s definition. Haidt calls the moral relativist-style response pattern (i.e. endorsing interference without universalising) an “enforceable-conventional” stance, which contrasts with “fully moralised” (endorsing interference and universalising) and “personal morality” (universalising without endorsing interference) stances, suggesting he really might not see it as any kind of
moralising at all (Haidt et al. [1993: 622]). Moral relativism may be false, but we ought not, by definitional fiat, refuse to count it as a moral view at all.

Having given their account of what moralising amounts to, Haidt et al. aimed to determine what kind of stance subjects would adopt toward the paradigmatic moral and conventional transgressions used in the moral/conventional task, as well as in response to scenarios describing disrespectful or disgusting actions.

Haidt’s study found that culture and SES did have an effect on the domain of morality. Specifically, Brazilian more so than North American subjects and low-SES more so than high-SES subjects moralised harmless but disgusting or disrespectful actions, such as cleaning one’s toilet with one’s national flag, eating one’s pet dog after it is accidentally killed by a stranger, and masturbating with the aid of a frozen chicken. Haidt takes his experimental work, along with the observations of cultural anthropologist Richard Shweder ([1991]; Shweder et al. [1997]), to support the conclusion that there are:

...cultural differences in the domain of morality. There does not appear to be a single list of content areas—even defined abstractly as harm, rights, and justice—that can capture the moral world of all peoples (Haidt et al. [1993: 625]).

It is time to step back and assess the bearing of Stich’s claims (against the background of experimental work like that of Haidt) on Joyce’s account of the evolution of morality and on the modified version of SSM I have suggested in this chapter. Joyce does focus on the fitness benefits of certain kinds of helpfulness in giving his evolutionary explanation of morality. While he is too cautious to rule out the possibility of moralities with other contents, he does say that concerns about interpersonal relations, especially involving harm and fairness, make up the bulk of “actual human morality” (2006: 66). Also, many of the signalling behaviours mentioned in the modified version of SSM I have proposed are helpful in nature (e.g. charitable giving, acts of heroism). If it turns out that helpfulness is but one small corner of morality,
explanations for the evolution of morality that focuses mostly on helpfulness might be thought importantly incomplete.

In what follows, I will discuss several strategies for defending a ‘Joyce-style’ approach to explaining the evolution of morality (and so defending the modified version of SSM I have suggested). These are the foot-stamping strategy, the redescription strategy, and the extend-the-strategy strategy. The upshot of my discussion is that such an approach is ultimately defensible, though much depends on empirical questions that are currently open.

Stich makes out that interpersonal relations might be only part (and a small part at that) of the content of morality. Haidt’s work can convey the same impression. One strategy in response is to insist that judgements (or norms, systems of rules, etc.) simply must have a certain content in order to count as moral judgements. Neil Levy can be read as adopting something like this strategy. Levy claims that moral systems:

...must be devoted, largely if not wholly, to concern for the welfare of other people [and] must systematise norms of justice and fairness... If it does not have the right kind of content, we should be reluctant to call any system of prescriptions a morality (2004: 44).

Judgements about purity, pollution, sacredness and profanity are, according to this strategy, just of the wrong kind to count as moral judgements. This is not to deny that what is eaten, what is worn, how the dead are treated and so on matter to many people, or to say caring deeply about such things is silly or primitive. It is, however, to exclude such concerns from the domain of morality proper. This might be termed the ‘foot-stamping’ strategy.

Adopting the foot-stamping strategy seems to me unwise at this stage in our understanding. Joyce is far from alone among philosophers in suggesting that the distinctive normative form of moral judgements lies in their putative inescapability and authority. Admittedly, studies in psychology
have not investigated judgements with exactly that normative form. Still, what psychologists have found makes it seem not unlikely that future work might turn up folks who judge that performing certain acts is self-polluting and wrong regardless of what one happens to believe, need or desire, and who take that supposed fact to give everyone reason to avoid such acts. It seems a live possibility that some folks might make judgements that have the (supposedly) distinctive normative form of moral judgements but that concern subject matter far removed from harm, fairness or any kind of interpersonal relations. I think that if one were to now stamp one’s foot about constraints on the content of morality, one would be in the same motion sticking one’s neck out for the empirical axe, which may not have far to fall.

An alternative strategy to simply denying that concerns about purity and pollution count as moral is to redescribe such concerns in terms of more familiar, interpersonal moral concerns, especially the concern with harming others. This seems to be Joyce’s favoured strategy for dealing with Stich’s challenge regarding the content of morality. Joyce warns against “assuming that what we in the West might count as a purely self-regarding action should be categorised as such in another culture” (2008: 253).

I call this the ‘redescription’ strategy. It opens with the observation that, although some experimental work has controlled for the effect of harm judgments on responses to disgusting transgressions, examples of supposed purity-based moralities are often drawn from broadly religious contexts. Haidt, for instance, supports his claim about cross-cultural variation in the moral domain by noting that “a great deal of the moral law of Judaism, Hinduism, Islam, and many traditional societies is explicitly concerned with regulating purity and pollution” (Haidt & Joseph [2004: 60]).

While gods (or ancestral spirits, as the case may be) are not typically seen as subject to brute physical harm, it is typically thought that such beings can be pleased or displeased. Violating a purity-based prohibition on eating certain foods might seem like a purely self-regarding violation from the standpoint of one who does not accept the religious framework within which
the prohibition is embedded. For a believer, though, who thinks such a transgression offends the gods (ancestors, etc.) by flouting their wishes, such a violation may not be seen as purely self-regarding at all. Purity-based rules like this may well be experienced as interpersonal, relating an agent not to other human persons but to supernatural persons (who, moreover, occupy particularly important positions in the agent’s worldview).

If the redescription strategy as executed above seems like too much of a stretch, consider the following variation. Notice that polluting transgressions are typically taken to have bad consequences (sometimes quite severe ones) for the transgressor’s community, should the proper purification rites not be performed. Viewed in this light, seemingly strange injunctions to avoid certain foods (for example) can be recognised as more familiar injunctions to avoid bringing harm to others. Purity-based rules might even be redescribed as straightforward prudential advice, since the transgressor himself is typically believed to suffer harm of some kind as a result of being polluted. Consider the following from anthropologist Mary Douglas, who studied the Yurok Indians of Northern California:

The Yurok so much believed that contact with women would destroy their powers of acquiring wealth that they held that women and money should never be brought into contact. Above all, it was felt to be fatal to future prosperity for a man to have sexual intercourse in the house where he kept his strings of shell money. In winter, when it was too cold to be out of doors, they seem to have abstained altogether... [It is] significant for understanding their idea of female pollution that for the Yurok men there was a real sense in which pursuit of wealth and of women were contradictory (2002/1966: 189-90).

One might worry that the redescription strategy does not take seriously enough the idea that purity and pollution as such are central to some cultures’ worldviews. Haidt would likely grant that actions that defile places, people, and objects are often believed to have harmful consequences but insist that, for participants in purity-based norm systems, the pollution caused by such acts figures as a deliberative consideration in its own right. The redescription

---

76 See Attridge (2004: 71-76) for examples drawn from many cultures of pollution-causing transgressions, the dangers such pollution was believed to pose, and the corresponding purification rites.
strategy, on this kind of view, just reflects educated, Western, liberal biases (Haidt [2007: 1001]).

Empirical work teasing apart considerations of purity and pollution from harms to self and others (human and otherwise) would be invaluable at this point. To my knowledge, though, such work has not yet been conducted, and the methodological challenges involved in carrying it out would be daunting. Consider: asking an Oriya man in India whether touching a menstruating woman would be wrong if doing so carried no risk of harm to self or others nor did it offend anyone (gods included) may be tantamount to asking him to entertain a conditional with an impossible antecedent.77 This problem worsens if we suppose impurity is seen not just as instrumentally bad but also as intrinsically harmful. Being an educated, liberal Westerner, I find it hard to get into the head of a participant in a purity-based norm system. For all I know, someone with that mindset may think that to be made impure (by the touch of a lower-caste individual, for instance) is to suffer a kind of harm.78 If this were so, it would make experimentally teasing concerns about purity apart from concerns about harm even harder.

The worries above aside, some concerns about purity and pollution should probably be expected to resist redescription in terms of harm or other interpersonal factors. After all, one of the striking findings of Haidt et al. (1993) was that actions judged to be harmless but disgusting triggered the same sort response as was triggered by the prototypical harmful transgressions used in the moral/conventional task. The next strategy I will discuss is not an alternative to the redescription strategy but rather meant to complement it.

What I call the ‘extend the strategy’ strategy concedes for argument’s sake that morality encompasses issues of purity and pollution that cannot be redescribed in terms of harm or other interpersonal concerns. This strategy seeks to explain moral thinking about these matters in the same kind of way

---

77 ‘If that were so,’ he may think, ‘who knows what would be right or wrong?!’
78 As Joyce notes (pers. comm.), it seems like the kind of thing one would wish on one’s enemies...
that Joyce explains thinking in moral terms about helpfulness. That is, the strategy seeks to establish that there are fitness benefits to behaving in the ways effectively motivated by moral judgements concerning purity and pollution. In essence, this strategy depends on treating Joyce’s focus on helpfulness as dispensable to an evolutionary explanation of the trait of making moral judgements. The evolution of that trait is to be explained by reference to its efficacy in motivating (by and large) fitness-enhancing behaviours, but if helpful behaviours turn out to be only one among several kinds of fitness-enhancing behaviours motivated by moral thinking, that does not invalidate the general approach.

Some care is needed in employing this strategy. It will not suffice to cite relatively modern examples like the Jewish kashrut (dietary rules) prohibiting (among other things) pork consumption together with controversial claims about the supposed health risks of eating pigs. It is not the current utility (in evolutionary terms) of specific norms about purity and pollution that is relevant, but the utility to our distant ancestors of having the capacity for moralising about such matters. Also, the evidence needed to substantiate claims about how making purity-based moral judgements provided fitness benefits to our distant ancestors is likely to be scattered throughout the literatures of numerous disciplines, if it is available at all. I will not develop this strategy in any great detail here. I will only indicate some of the issues that would need to be confronted in doing so.

79 See Simoons (1994) for a critical discussion of the trichinosis hypothesis about Hebrew pork taboos.

80 Although, see Ogbeide (1974) for an excellent example of the kind of empirical work needed to assess the fitness benefits or drawbacks of purity-based norms (this study reports the negative nutritional impact of food taboos on children and pregnant women in traditional Nigerian societies).

81 A worry that I think can be fairly swiftly set aside is that the ‘extend the strategy’ strategy is culturally insensitive insofar as it tries to fit diverse cultures’ beliefs and practices into a mould shaped by a single culture’s point of view (Western culture is again the culprit). Mary Douglas, for instance, decries the “medical materialism” that informs many anthropologists’ investigations of other cultures, saying that:

…it is one thing to point out the side-benefits of ritual actions, and another thing to be content with using the by-products as a sufficient explanation. Even if some of Moses’s dietary rules were hygienically beneficial it is a pity to treat him as an enlightened public health administrator rather than as a spiritual leader (2002/1966: 30).

However, what matters as far as the strategy I am discussing is concerned is not how participants in purity-based moralities understand or justify their rules against polluting acts.
One may worry that even if some supposed pollution-avoidance is susceptible to evolutionary explanation in the way just sketched, much of it seems to carry no fitness benefits or to be actively detrimental to one’s reproductive prospects. For example, recall the Yurok belief that bringing women and money into contact or even proximity can destroy one’s ability to accumulate wealth. Judging by the distribution of tribesmembers’ birthdays throughout the year, this belief is apparently sufficient to make many Yurok men abstain from sex during the winter months when they, their wives, and their money are all trapped inside together (Douglas [2002/1966: 189]). A purity-based moral judgement that motivates one to go months without sex seems unlikely to advance one’s reproductive fitness. I will not multiply examples here but instead rest content with the assertion that a trawl through the purity-based prohibitions and requirements of various cultures would net many examples that support this worry.

One may also worry that where supposed pollution-avoidance does plausibly secure some fitness benefit, it is hard to see how one’s motivation to shun supposedly polluting thing(s) would need bolstering (or at least, harder than it is to see how one’s motivations to help others might need it). For example, some of the most promising cases for the the extension strategy are certain food taboos, rules about ritual washing, and regulations governing disposal of the dead. Preserving personal and local purity in such cases may incidentally secure significant health benefits. However, one’s motivations to clean up any corpses lying about, for instance, or to avoid the sacred mushrooms that make the shaman spasm and froth, are probably fairly robust. Strong emotional responses such as disgust and fear are effective motivators that require little in the way of bolstering from clouty moral judgements.

What matters are the effects of people’s purity-based beliefs on their behaviour and thus on their reproductive fitness. The side-benefits or by-products of purity-based moral systems are precisely what drive the kind of explanation I am considering. But adopting this explanatory strategy does not require claiming that what people really care about when it comes to purity and pollution is staying healthy rather than pleasing God (the gods, the ancestors, etc), so Douglas’ worry does not apply.
I take the two worries just laid out to present the extend the strategy with something of a dilemma. Some remarks can be made to blunt the horns, though.

First, the success of the strategy does not depend on every purity-based norm having a plausible adaptive rationale, just as the success of Joyce’s initial case does not depend on every morally motivated instance of helpfulness being fitness-enhancing. A key question as far as this strategy is concerned is just what adaptive benefit might accrue to having a faculty for moralising that is sensitive to considerations of purity and pollution. That is a currently open empirical question.

Also on this point, where one looks for a fitness benefit of purity-based moralising is crucial. I have so far been considering the possibility that making purity-based moral judgements might benefit individuals. The fitness benefits of purity-based moralising may also (or mostly) accrue to cultural groups. Shared purity norms are important markers of group identity that may strengthen group cohesion (Bowie [2000: 73]; Meyer-Rochow [2009]). While considerable empirical detail would need to be filled in to make the extend the strategy work, the prospects for explaining purity-based moralising in terms of its adaptive benefit is not as unpromising as the first worry above may suggest.

Second, while the motivational boost provided by purity-based moralising may well be superfluous when supposedly polluting things trigger strong emotional responses, such responses need not always be triggered by such things. Sticking with the example of food taboos, Begossi et al. (2004) report that among certain Amazonian tribes, piscivorous (i.e. fish-eating) fish are a taboo food for pregnant women. Such fish are not in and of themselves particularly repulsive. They may even be tasty. As it turns out, however, these fish contain relatively high concentrations of toxins and contaminants, in virtue of being near the top of their food chain (Begossi et al. [2004: 1341]). In cases like this, the motivation to avoid forbidden foods may well sometimes need boosting. If judging in moral terms achieves this, then doing so would
help secure the benefits (in this case health benefits) of refraining from eating forbidden foods.

Also on this point, it seems worth mentioning that disgust responses and moral judgements are not entirely separate; the latter can help create the former. Suppose a fundamentalist Christian happened upon two men holding hands outside a hospital. Imagine she learned that the two men are friends, one of whom is recently bereaved. Now imagine instead that she learned the two men are lovers awaiting the decanting of their first cloned child. I am willing to bet that in the second case, a strong disgust reaction would follow hard on the heels of the Christian’s realisation that she had happened upon homosexuals playing God. Insofar as moralising actions can create disgust responses (or prompt other strong emotions), the second worry raised above may be less damaging to the extend the strategy strategy than it seems.

There is one more reply I think can be made in response to Stich’s comments about the extent of the moral domain and the importance of helpfulness in a Joyce-style explanation for the evolution of morality. This final reply is less of an argumentative strategy and more of a straightforward observation: call it the ‘origin vs. elaboration’ point. It is important to note that the claim Stich attributes to Joyce – that the selective pressures favouring the capacity to make moral judgments sprang mostly from the fitness benefits of certain kinds of helpfulness – is compatible with there being many norms having nothing to do with interpersonal relations, even granting that such norms are genuinely moral norms. That claim concerns the initial driving forces in the evolution of morality. Making it does not commit one to the claim that the domain(s) of action it was fitness-enhancing for our distant ancestors to judge in moral terms are co-extensive with those over which the human capacity for moral judgement is now exercised. Of course, giving an account of how and why the moral domain expanded would not be trivial. The bare logical point stands, though: even indisputable evidence of cross-cultural differences in the domain of the moral would not sink a Joyce-style account of the evolution of morality.
2.4. Conclusion

This chapter has been concerned with the issue of how best to specify the explanatory target of an evolutionary account of morality. I have surveyed a range of work done under the rubric of ‘evolutionary explanation of morality.’ I have divided this work into attempts to explain specific kinds of behaviour, attempts to explain certain social or group-level phenomena, and attempts to explain some element of human psychology. Much of this work is interesting and important in its own right but does not aim at the right explanatory target qua evolutionary explanation of morality. I have argued (following Richard Joyce) that the appropriate explanatory target of evolutionary accounts of morality is the tendency to make moral judgements. We want to know why we think in moral terms. My favored answer, in short, is that moral thinking is motivationally efficacious thinking.

I have proposed a revised version of SSM that marries costly signalling theory and the theory of sexual selection to an account of the evolution of morality according to which we have been selected to make moral judgements because of the motivational efficacy of so judging. In sum, my revised version of SSM claims that sexual selection for certain kinds of signalling behaviours could have also generated selection for the tendency to make moral judgements, since making such judgements makes acting in those sexually-attractive ways more likely.

Finally, I have defended my favored approach to explaining the evolution of morality against a pair of objections that are ably pressed by Stephen Stich.

The first of these objections concerns the supposed universality of moral judgments (as I and Joyce conceive of them). I have responded to this objection by undermining the empirical evidence – an attack on the moral/conventional distinction – upon which the objection is based. In brief, there is an alternative interpretation of the results of this attack that leaves the moral/conventional distinction intact.
The second of Stich’s objections concerns the extent of the moral domain, which he (along with Jonathan Haidt) believes encompasses more than the considerations of harm and fairness central to Joyce’s account of morality. My response to this objection has been to outline several strategies for reconciling Stich’s observations with my (and Joyce’s) favoured approach to explaining the evolution of morality. While much in the way of empirical detail remains to be filled in, the strategies outlined – especially the redescription and extension strategies – show how this approach can be plausibly maintained.
3.1. Introduction

This chapter takes up the issue, raised in Chapter 1 (see §1.2.2b), of how to situate sexual selection in the context of other processes often appealed to in discussions of human evolution. In Miller’s initial presentation of SSM, these other processes are given short shrift. Miller describes kin selection and reciprocal altruism, asserts their failure to account for helpfulness toward unrelated unlikely reciprocators, then moves swiftly on to make his case for the importance of sexual selection in understanding the evolution of such behaviour (2000: 297-304). In effect, Miller initially offered an incomplete argument from elimination for the importance of sexual selection. This too-cursory treatment of processes other than sexual selection is improved upon in Miller’s updated version of SSM. Even there, though, Miller maintains that sexual selection is a special kind of evolutionary process distinct from natural selection (2007: 98).

My first goal in this chapter is to give more careful consideration to what processes other than sexual selection can contribute to evolutionary explanations of the capacity to make moral judgements. Recall, the modified version of SSM I presented in Chapter 2 (see 2.2.3b) claims that the capacity to make moral judgements evolved because such judgements effectively motivate behaviour favoured by inter-sexual selection. The signalling behaviours in question are, by and large, helpful behaviours. Defending SSM thus stated will require showing that other processes taken to be relevant in the evolution of helpfulness leave explanatory space for sexual selection to fill.

Kin selection, reciprocity, and mutualism form a familiar litany to those working on the evolution of morality. I will not spend too much time rehearsing work already done (just long enough to make clear that these processes have more to contribute than Miller makes out). I will devote far more attention to a relatively neglected kind of view about the evolution of
morality, namely, the by-product explanation. Jesse Prinz has recently championed the by-product explanation ([2007a]; [2008a]; [2008b]; [2009]). According to Prinz, our capacity to make moral judgements is an evolutionary accident, a “by-product of capacities that were evolved for other purposes” (2009: 168). If Prinz’s by-product explanation is correct, then seeking some adaptive explanation for our capacity to make moral judgements (SSM, for instance) is a misguided endeavour. I will present Prinz’s case for the by-product explanation. The crux of my response to Prinz is a critique of his use of the concept of innateness.

My second goal in this chapter is to assess the claim that sexual selection is a somehow special kind of evolutionary process distinct from natural selection. I will begin by laying out Darwin’s distinctions. I will then draw on the biological literature to find or reconstruct arguments for either collapsing or maintaining Darwin’s distinctions. While biologists’ views exhibit considerable disagreement about the nature of the relationship between natural and sexual selection, there seems to be precious little attention devoted specifically to the issue of how to distinguish evolutionary processes. Accordingly, I will attempt to lay out a criterion by which to distinguish evolutionary processes. Lastly, I will present and evaluate Amotz Zahavi’s ([1997]; [2007]) distinction between “utilitarian selection” and “signal selection.” Zahavi does not defend exactly Darwin’s distinction, but his work might be pressed into service to defend Miller’s claim that inter-sexual selection is importantly different to kin selection, reciprocity, and the like. I will argue that Zahavi’s distinction is, ultimately, unsustainable.

My final aim in this chapter is to consider the consequences for SSM of the collapse of Darwin’s distinction between natural and sexual selection. SSM, I argue, would not be vitiated by the finding that there is no sustainable distinction to be drawn between natural and sexual selection. SSM remains an interestingly novel view about the evolution of morality even if sexual selection is best subsumed under natural selection.
3.1.1. Evolutionary Explanations for Helpfulness

In Chapter 2, I offered a version of SSM according to which the capacity to make moral judgements evolved because such judgements effectively motivate signalling behaviours favoured by intersexual selection. Many of the signalling behaviours in question are helpful ones. Also in Chapter 2, I noted that many of the helpful behaviours discussed in evolutionary biology seem at least at first glance to be fitness-sacrificing as well: food sharing and alarm calling behaviours, for instance, appear to advance the fitness of their beneficiaries at some cost to the helper. The helpful signalling behaviours that feature in SSM also fit this pattern.82

There are two tactics that can be adopted to explain apparently fitness-sacrificing helpfulness. One is to show that such helpfulness is not actually fitness-sacrificing for the helper, by identifying some fitness benefit the helper gains by helping. The other is to show how the existence of such helpfulness is consistent with an evolutionary perspective on behaviour despite it being costly in terms of fitness for the helper.

In what follows, I will present several explanations for helpfulness that take up one or the other of these two options. Much can be done to explain apparently fitness-sacrificing helpfulness without appealing to sexual selection. Even so, an appeal to sexual selection is not otiose. When it comes to explaining the evolution of helpfulness, there is explanatory work left for sexual selection to do.

3.1.1a. Mutualism, Reciprocity, and Kin Selection

The first and most straightforward explanation for the evolution of helpfulness that I will consider is mutualism, by which I mean explanation in terms of the

82 While a rigorous treatment of costs and signalling will have to wait until Chapter 4 (see §4.2.1), for now it will do to observe that an agent acting generously, heroically, fairly, and so on can be plausibly taken to provide fitness benefits to others at some (perhaps considerable) cost to herself.
immediate benefits the helper gains by helping.\textsuperscript{83} The benefits gained by helper and beneficiary need not be equal in order for this kind of explanation to work; it is enough that the helper be better off helping than not. When helping is to the advantage of all involved, there is no great mystery as to why helpfulness would have evolved, though there may be many interesting proximate-level questions about how helpfulness like this originates and is maintained.

Examples abound of helpfulness explicable in terms of mutualism. To take a well-known case, bacteria of the genus \textit{Rhizobium} live in the roots of leguminous plants like peas and beans, using nitrogen from the soil to build organic compounds for their host while their host sends them nutrients and energy. Pack hunting has also been explained in terms of mutualism. For instance, African wild dogs (\textit{Lycaon pictus}) apparently enjoy the greatest hunting success when in the largest packs, suggesting that the individual dogs help each other to make and defend a kill and all reap benefits by doing so (Creel & Creel [1995]).\textsuperscript{84} Alarm calling and mobbing defence in birds are two more kinds of helpfulness that may plausibly be accounted for in terms of mutual benefits (Dugatkin [1997: 85-89]).

Not all helpfulness brings an immediate benefit to the helper. Robert Trivers’ (1971) theory of reciprocal altruism did much to account for apparently fitness-sacrificing helpfulness that does not bring an immediate benefit to the helper. Trivers realised that such behaviour could be fitness-advancing for the helper so long as certain conditions were met. First, there must be an asymmetry between the cost and the benefit of helping, with the cost of helping being small compared to the benefit helping confers. Second, there must be the potential for individuals to interact repeatedly and to switch roles (i.e. helper vs. beneficiary). Third, individuals must have the ability to

\textsuperscript{83} ‘Immediate’ here helps distinguish mutualism from reciprocity, to be discussed below. Herre et al. (1999) review research on mutualism and describe in detail some fascinating examples of mutualistic partnerships.

\textsuperscript{84} Finding the right currency in which to calculate benefits is a big challenge here. Creel & Creel ([1995: 1337]; see also [2002]) settle on kilograms of meat per dog per kilometre travelled during hunting, on the grounds that this measure best incorporates the costs of hunting.
recognise those who have helped them in the past. Fourth and finally, individuals must have a tendency to help those who have helped them in the past and to withhold help from those who have failed to help them. Under these conditions, the immediate cost of helping another individual can be recouped (and more) over time as one’s beneficiaries repay one’s helpfulness in kind. Helpfulness under these conditions can thus be to a helper’s long-term advantage.

Food sharing among vampire bats (*Desmodus rotundus*) is typically taken as an example of reciprocity in action (Wilkinson [1984]).\textsuperscript{85} Vampire bats that have recently fed will share food with hungry bats by regurgitating some of their last blood meal. While much of the food sharing among vampire bats occurs between close kin (mothers and offspring) and so is explicable in terms of kin selection, there remain instances of food sharing between bats that are only distantly related (Wilkinson [1984: 182]).

As it turns out, vampire bats satisfy the conditions for the evolution of helpfulness via what Trivers called reciprocal altruism. They live in stable social groups and so have the chance to interact repeatedly. They experience quite variable hunting success, so individuals have the chance to alternate roles in interactions with others (i.e. donor vs. recipient of blood). Furthermore, individual bats are recognisable by their distinctive vocalisations. Crucially, the cost to a well-fed bat of donating blood is small relative to the benefit conferred on a hungry bat by doing so.\textsuperscript{86} In support of an

---

\textsuperscript{85} I say ‘typically’ because this canonical example has been questioned. Stevens *et al.* ([2005]; see also Stevens & Hauser [2004]) point out the marked rarity of non-kin as compared to kin blood donation and suggest that what Wilkinson saw as reciprocity might instead be the result of recognition error. More generally, Stevens and colleagues believe that:

...cognitive constraints on temporal discounting, numerical discrimination, learning and memory, and other components, limit the ability of many species to implement and maintain reciprocally altruistic strategies (Stevens & Hauser [2004: 64]).

Whether there are good examples of animal reciprocity is, for my purposes, a side-issue. The animal cases described here are meant to illustrate the general kinds of explanations for helpfulness I am considering. Reciprocity is undeniably important in the case of human helpfulness, which is my focus here.

\textsuperscript{86} Wilkinson reported (1984: 183) that the bats store and metabolize blood meals in such a way that, when a bat is full or close to it, a given volume of blood represents fewer hours of life than when a bat is close to empty. That is, were full bat $A$ to donate $x$ units of blood to
explanation in terms of reciprocity for vampire bats’ apparently fitness sacrificing helpfulness, Wilkinson found that non-kin food sharing occurred only between frequent roost mates.

Trivers’ proposed explanation for apparently fitness sacrificing helpfulness has come to be called ‘direct’ reciprocity, since it involves pairs of individuals that trade favours with each other. Richard Alexander extended Trivers’ work by suggesting that reciprocity might also be indirect:

[In indirect reciprocity the return is expected from someone other than the recipient of the beneficence. This return may come from essentially any individual or collection of individuals in the group. Indirect reciprocity involves reputation and status, and results in everyone in a social group continually being assessed and reassessed by interactants, past and potential, on the basis of their interactions with others (1987: 85).

Alexander’s proposal has received both theoretical and empirical support. Nowak & Sigmund (1998) offered mathematical models and computer population simulations to show that helpfulness can evolve if helping is conditional on the beneficiary’s “image score” being above a certain threshold (which varies depending on the costs and benefits of helping). Bshary & Grutter ([2006]; see also Bshary & D’Souza [2005]) report field observations and the results of laboratory work to show that relationships between cleaner fish (Labroides dimidiatus) and their many client species are maintained by indirect reciprocity.87

There may be a temptation to see signaling to potential mates via costly helpfulness as a case of trading helpfulness for sex, and to class it as an

\[A \text{ would be } y \text{ hours closer to death by starvation but } B \text{ would gain a } z \text{ hour reprieve from death by starvation, where } z > y.\]

87 Cleaner fish were Trivers’ original example of reciprocal altruism. Trivers thought that clients refrained from eating cleaners after a service so they could return to the same cleaning station and be serviced by the same cleaner again in future. Bshary and colleagues’ work reveals several features of cleaner-client interaction that seem more amenable to explanation in terms of indirect reciprocity. Clients appear to interact with a particular cleaner or not based just on observation of that cleaner’s last interaction with a client. Also, Bshary & Grutter (2006: 976) show that cleaner behaviour evinces an audience effect: cleaners cleaning clients in the presence of other potential clients are less likely to ignore their current client’s parasites in favour of eating its mucus instead (which cleaner fish apparently prefer, according to Grutter & Bshary [2003]). As Bshary notes – and apropos of footnote 4 above – indirect reciprocity via image scoring is less cognitively demanding than the kind of direct reciprocity between cleaners and clients envisioned by Trivers.
instance of reciprocity. While it is true that reciprocity does not require repayment to come in the same currency as was initially tendered, notice that it is possible for costly signalling to work under conditions that do not allow for reciprocity (cf. Roberts [1998: 430]). As described above, for reciprocity to work, there must be a cost/benefit asymmetry to helping, such that the benefit bestowed upon the receiver is greater than the cost incurred by the helper. This need not be the case for helping to work as a costly signal. There, it only need be the case that the benefit eventually bestowed on the helper outweighs the cost of helping in first place. Notice also that bestowing a benefit on the helper may not be costly at all for the receiver of the help (or for those who observe the help being provided, in the case of indirect reciprocity), and doing so may in fact be beneficial. This is plausibly the case when the benefit is mating and the helper has, by helping, proven themselves a high-quality mate. In the specific case of costly signalling via helpfulness to potential mates, it seems, there is no incentive for the receiver of the help to defect. This makes the costly signalling scenario even more different to cases of reciprocity.

Moving on, it is probably apparent that mutualism and reciprocity seem ill-suited to explaining helpfulness when helping does not bring either an immediate or a delayed benefit to the helper. William Hamilton (1964) solved large sections of the puzzle posed by this kind of apparently fitness-sacrificing helpfulness when he conceived of the pieces not as individual organisms but

---

88 At least, not as Trivers conceived of it. More liberal understandings of 'reciprocity' may license calling some kinds of costly signaling cases of reciprocity. Doing so strikes me as strange; even though both parties benefit, one party (i.e. the signal receiver) may never pay a cost. I am content here to show how costly signalling differs from reciprocity as Trivers understood it, since my aim is not not to judge between different definitions of 'reciprocity.' For a useful discussion of that issue, though, see Rothstein & Pierotti (1988).

89 To make this point less abstract, just suppose Hawkes' show-off hypothesis is correct. The benefit to a woman of having some meat is probably quite small compared to the costs paid by the hunter who obtained that meat. Here, the cost/benefit asymmetry is the reverse of what is required by reciprocity. Nevertheless, paying those costs means the hunter impresses the woman enough for her to decide to have a child with him. That benefit is probably greater than the costs of hunting. The asymmetry here means costly signaling is at least a possibility. I have put this down in a footnote because the empirical details are uncertain and I do not want that uncertainty to detract from the more formal point made above: costly signaling via helpfulness can work when reciprocity cannot.
as much smaller units, namely, genes. Apparently fitness-sacrificing helpfulness could actually be fitness-advancing, Hamilton realised, even if it was disastrously harmful to the helper, so long as the helper and those helped were sufficiently closely related.

The more closely two individuals are related, the more likely it is they share genes that are identical by descent, meaning copies of the same ancestral gene. Thus, Hamilton claimed, helping kin may evolve despite decreasing the helper’s “personal” fitness provided it generates a compensating increases in the helper’s “inclusive fitness” (1964: 8). That is to say, if the cost of helping decreases the number of genes the helper contributes to the next generation in the form of its own offspring by less than the help increases the number of genes (identical by descent to the helper’s) passed on by the helper’s relatives, then helping makes evolutionary sense. This inequality has come to be known as Hamilton’s Rule, and this way of accounting for the evolution of apparently fitness sacrificing helpfulness, as kin selection.

Insects of the order Hymenoptera, such as certain ants, bees, and wasps, have furnished some of the most puzzling examples of apparently fitness-sacrificing helpfulness. These species are eusocial, having sterile female worker and/or soldier castes that spend (and often expend) their lives feeding and protecting their queen and her multitudinous offspring, entirely forgoing reproduction of their own. It thus seems fitting that these organisms provided quite striking support for Hamilton’s ideas about kin selection. All Hymenopteran species are haplodiploid: males develop from unfertilised eggs

---

90 Although Hamilton deservedly gets the credit for the theory of kin selection, the crucial insight was not original to him. Ronald Fisher (1914: 315) gave the following example in an early paper on eugenics:

A is the eldest son, and stays at home; his brother B goes to the wars; then so long as A has some eight children, it does not matter, genetically, if B gets killed, or dies childless, there will be nephews to fill his place.

And, of course, J.B.S. Haldane famously joked that he would not sacrifice himself for his brother but would do so to save eight cousins (quoted in McElreath & Boyd [2007: 82]).

91 Hamilton’s Rule is sometimes summarized as saying that helpfulness can evolve if $rb > c$, where $r$ is the relatedness of helper and beneficiary, $b$ is the benefit bestowed on the beneficiary, and $c$ is the cost of helping.

92 Darwin wrote of sterile worker and soldier ants – apparently fitness-sacrificing helpers par excellence – that they were “one special difficulty, which at first appeared to me insuperable, and actually fatal to my whole theory” (1998/1859: 192).
and have only one set of chromosomes (haploid), while females develop from fertilised eggs and have two sets (diploid). This genetic arrangement means that Hymenopteran females share, on average, half their genes with their offspring but three quarters of their genes with their sisters. So, the worker daughters of a hive queen can actually do better, in terms of fitness, by helping their queen produce more daughters than they would do by producing offspring of their own (Bourke & Franks [1995: 77-82]). Hamilton took the fact of haplodiploidy and the idea of kin selection together to account for the prevalence of eusociality in Hymenopterans and the relative dearth of it outside that order.

A couple of cautions are in order here. First, species need not be haplodiploid in order for kin selection to lead to the evolution of helpfulness. The important thing is just that the inequality described by Hamilton's Rule holds, and there are several ways in which high degrees of relatedness between interacting individuals might come about. An example of eusociality due not to haplodiploidy but to frequent inbreeding comes in the unlovely form of the naked mole rat (mentioned in §1.1.2a). Limited dispersal can also lead to the evolution of kin-selected helpfulness. Cooperative breeding provides many examples of helpfulness like this. For instance, until Seychelles warblers (*Acrocephalis sechellensis*) can secure breeding territory of their own they help their parents breed, defending and maintaining the family nest and feeding newly-hatched young. However, helpers apportion their help according to their relatedness to the chicks, feeding half-siblings significantly less than full siblings (Komdeur [1994]).

Second, the theory of kin selection should not lead us to suppose that kin will always help each other. Again, the important thing is that the inequality described by Hamilton's Rule holds, and under some circumstances the costs of helping kin can be prohibitive, no matter how close the relationship or how great the benefit conferred. For example, consider West *et al.* (2001) on fig wasps. Wingless male fig wasps are born, develop, and must find a mate inside a single fruit. When only one or a few females lay eggs in a fruit, the males within it are, on average, closely related to each other. When
many females lay in the same fruit, the males within are largely unrelated to each other. Surprisingly, though, the average relatedness of males within a fruit has no effect on the fierceness of male-male competition within that fruit. The factor that determines male competitiveness is the availability of mates. When females are abundant, male conflict is relatively forgiving, but when females are scarce, males are ruthless. In the latter case, the benefits of helping siblings by restraining oneself in the competition for mates do not outweigh the costs of doing so: if only one of you can reproduce, then sons and daughters are better than nephews and nieces.\textsuperscript{93}

3.1.1b. Human Helpfulness and Mistaken Identity

The explanations for helpfulness presented in the previous section seem able to explain some but not all human helpfulness. Mutualism can certainly account for some human helpfulness. For example, early humans engaged in the equivalents of pack hunting, alarm calling, and group territorial defence. Obviously, though, much human helpfulness cannot be explained in terms of mutualism. Rendering aid to the feeble, for example, provides no immediate benefit to the helper. Thus, mutualism is at most part of the explanation for human helpfulness. Kin selection is undeniably important in explaining the evolution of human helpfulness. Since early humans lived in smallish groups composed of several extended families (Boehm [2001]), it is plausible to suppose that helping to feed, defend, and generally promote the well-being of one’s fellow group members often netted helpers inclusive fitness benefits. However, not all human helpfulness (now or during our evolutionary history) involves individuals sufficiently closely related that helping pays off in this

\textsuperscript{93} Much more could be said about the complexities involved in kin selection. The very name itself is perhaps misleading, since interactions between close relatives who share many genes identical by descent are not the only contexts in which kin selection is important. If a particular gene (or gene complex) caused its bearers to display a distinctive trait, to recognise that trait in others, and to give preferential treatment to those others, then that gene could spread by ‘kin’ selection operating on largely unrelated individuals. This so-called ‘greenbeard’ effect (Dawkins [1989/1976]: 89) has received some empirical support. Keller & Ross (1998) reported that in fire ants (*Solenopsis invicta*), workers who are heterozygous Bb at a particular gene locus can tell homozygous BB queens from heterozygous Bb queens, since only queens with a b allele produce a certain pheromone. Bb workers kill any BB queens that migrate into their nest. As Keller & Ross write, “selective elimination of BB queens results in an overall increase in the reproductive success of Bb queens and thus in the number of copies of the b allele transmitted to subsequent generations” (1998: 574).
way. Similar things can be said regarding reciprocity. Reciprocity, direct and indirect, plausibly explains much of human helpfulness. As Trivers noted, our ancestors lived under relatively stable social conditions that were conducive to reciprocal exchanges of resources (e.g. food, tools), knowledge, and physical aid (e.g. rescuing from danger, protecting against threat). Furthermore, humans have the cognitive equipment needed to manage such exchanges. Trivers even suggested (1971: 49-50) that moralistic aggression, gratitude, and guilt are evolved psychological mechanisms for maintaining reciprocal arrangements in the face of would-be cheaters. However, not all human helpfulness occurs under conditions allowing for direct reciprocity or for the kind of reputation-boosting useful in indirect reciprocity. Nor does reciprocity-based helping scale up well from extended family groups to larger and more complex societies.

The implications of the fact that human helpfulness extends beyond close kin or those likely to reciprocate are far from straightforward. Miller (2000: 300, 303-304) takes this fact to imply that kin selection and reciprocity are only part of the explanation for human helpfulness. The missing piece of the puzzle, he thinks, is sexual selection. Miller’s inference may be too hasty, though. It is possible that some (perhaps much) of the helpfulness toward unrelated unlikely reciprocators Miller seeks to explain in terms of sexual selection is simply, in evolutionary terms, a mistake on the part of the helper.

The mistaken identity explanation for helpfulness is nothing new. Some instances of helpfulness have been explained as mistakes by the helper. To take a well-known example, certain birds sometimes raise cuckoo chicks instead of their own (Davies & Brooke [1989]). Cuckoo chicks receive help, in the form of food and protection, from their foster parent(s), whose behaviour may well be fitness-sacrificing given that raising a cuckoo chick reduces the resources available to raise chicks of their own. Foster parents’ helpfulness, though, results from the limitations under which they must operate. Their need to forage means they cannot constantly guard their nests against cuckoo

---

But things may not be so simple – more on this point shortly.
infiltration. Accomplished egg-mimicry by cuckoos means they cannot infallibly distinguish between their own eggs and cuckoo eggs.

One might wonder why mistakes continue to be made in populations of organisms undergoing natural selection. One might think that individuals less prone to error would enjoy a selective advantage over, and eventually replace, more error-prone individuals. Of course, the benefits of a reduced error rate must be balanced against the costs of building machinery able to make finer-grained discriminations. The extent to which errors will be weeded out by selection depends on the interplay between error costs and the costs (and even the availability) of more accurate machinery. Thus, in some cases, organisms might be better off (or stuck with) error-prone detection systems.95

Returning to the human case, helpfulness toward non-kin unlikely to reciprocate might be explained as follows. Ancestral social environments contained mostly kin, potential reciprocators, and folk who would spread word of one’s misdeeds if they observed any. In such an environment, a tendency to help the individuals one lives among and frequently interacts with would be favoured by kin selection and reciprocity. Having an actual ‘kin detector’, so to speak, would seem unnecessary, even if something like that could be built by selection. The same might be said of a faculty for calculating the costs and benefits of helpfulness versus selfishness on a case-by-case basis. But while a tendency to help those around one nets fitness benefits in ancestral environments, such a tendency misfires in modern social settings. Nowadays, most of us interact mostly with non-kin, while anonymous, one-off social encounters are common. So, our evolved tendencies toward helpfulness might now lead us to act in ways that bring no reward via kin selection or reciprocity.

To take stock: there are explanations available for helpfulness that identify the evolutionary benefits of such behaviour, at least in certain contexts. These explanations can account for much of human helpfulness.

95 And in the cuckoo case, a deception versus detection arms race complicates matters even further.
Miller is right to point out that these explanations alone cannot account for all human helpfulness. Specifically, helpfulness toward unrelated, unlikely reciprocators remains puzzling. Where Miller goes wrong is in immediately reaching for another kind of explanation that depends on revealing hidden benefits to helpfulness (i.e. SSM). We must recognise the possibility that a quite different strategy could suffice to explain the puzzling leftover helpfulness. I take the mistaken identity explanation outlined above to be a live option. Where, then, does this leave Miller’s project (and my own) of elaborating and defending SSM? Why appeal to sexual selection to explain the benefits of helpfulness and the advantages of making judgements that effectively motivate it, when one could talk just in terms of kin selection and reciprocity, explaining away leftover helpfulness as error?

On the one hand, it would be unfortunate to neglect a useful weapon in the evolutionist’s explanatory arsenal. Some helpfulness escapes the sights of kin selection and reciprocity, and if we can zero in on it using sexual selection, well and good. On the other hand, we should not be trigger-happy. Over-enthusiastic applications of evolutionary theory to human behaviour are all too common. In the end, an appeal to sexual selection to explain elements of human helpfulness will be justified (or not) depending on the quality of the evidence on offer. In Chapter 5, I will consider what empirical support can be mustered for SSM. For now, I am content if my discussion has situated sexual selection more carefully among the various other ways of explaining the evolution of human helpfulness than did Miller’s initial presentation of SSM.

3.2. The By-Product Explanation

In this section, I will discuss a view of human morality that stands in competition with the revised version of SSM I proposed in Chapter 2. Recall, that version of SSM holds that the trait of making moral judgements evolved because such judgements effectively motivated helpful behaviour that functioned as a signal in inter-sexual selection. In the previous section, I discussed the mistaken identity idea, which raised questions about the need to appeal to sexual selection in accounting for human helpfulness. It is worth
stressing at this point that the mistaken identity idea is entirely compatible with the general approach to the evolution of morality taken by Ruse, Joyce, and myself. The idea I will discuss in this section, by contrast, stands in opposition to any explanation of our tendency to make moral judgements that appeals to some supposed adaptive advantage of doing so. The so-called ‘by product’ explanation claims that the fact that we make moral judgements is an evolutionary accident. If the by-product explanation is right, then the approach to the evolution of morality taken by Ruse, Joyce and myself is badly off-track. I turn now to consider this explanation.

Jesse Prinz has recently championed the by-product explanation in a series of publications, claiming that human morality is “an accident”, the “by-product of capacities that were evolved for other purposes” (2009: 168). Prinz sets up his view in opposition to Joyce- and Ruse-style views that “morality is an evolved capacity” (2009: 168). The following quote makes clear that Prinz takes himself to be offering an alternative to such views:

[Our] capacity to moralise could be an evolved adaptation that occurred after we split from the ancestor that we share with chimpanzees. Or, alternatively, the capacity to moralise could be a by-product of other capacities, which evolved for other purposes (2007: 263).

Prinz favours the latter alternative. I will quote Prinz at some length here, since responding to his claims is largely, I think, a matter of clarifying material too densely and swiftly presented. Prinz writes:

It seems...that we have good evidence for the claim that morality is an evolved capacity. Animals may not have moral systems in exactly the same sense that we do, but the resemblance is intriguing. It is tempting to conclude that human morality is an evolutionary successor to capacities found in other species. On the [sic] picture, morality is innate. We are born to be good. The concept of innateness is closely related to domain specificity. To say that a capacity is innate is, in part, to say that we have biological machinery dedicated to attainment of that capacity. Friends of innateness claims often emphasise universals. If morality is part of the bioprogram, and the evolved aspects of human psychology are generally monomorphic, then there must be moral universals, i.e. there must be aspects of our moral systems that are found amongst all normally developing members of our species (2009: 168).

96 Much the same material is presented in Prinz’s several publications on this topic. I will refer mostly to his latest publication, but that one is representative of the others.
Prinz goes on to say that the claim ‘there are moral universals’ can be interpreted in three different ways (2009: 168). Firstly, “immodest moral nativism” says that the human moral universals are moral rules with specific, fixed contents. “Modest moral nativism” says that all normal humans have a “morality acquisition device,” which consists of schematic moral rules and which takes inputs from culture to fill out those rules. Finally, “minimal moral nativism” says that all normal humans have a “moralisation mechanism,” meaning a faculty which “converts” nonmoral rules to moral rules. Prinz rejects all three versions of the claim that there are human moral universals. He takes this to support the conclusion that “morality...is a byproduct of capacities that were evolved for other purposes. Morality is a spandrel. There is no mechanism dedicated to the acquisition of moral norms” (2009: 168).

Prinz thinks our capacity for morality is a by-product of our emotional capacities and our capacities for rule-formation, memory, imitation, and mind-reading (2007: 270-272). To get into the substance of Prinz’s positive view would take us too far afield from our immediate concern with SSM as an account of the evolution of morality and the by-product explanation as a rival to that account. It is the overall structure of Prinz’s argument for the by-product explanation with which I want to take issue.

Talk of innateness, nativism, universality, domain specificity, dedicated machinery, by-products, and spandrels comes thick and fast in Prinz’s argument for the by-product explanation. It is hard to get clear on how what he says cuts against the claim that our capacity to make moral judgements is an adaptation. In the end, I think, it does not. My diagnosis of the disagreement between Prinz and the likes of Joyce and Ruse is

97 Regarding the immodest claim, rules against harming others and rules against incest are two of the candidate universals Prinz considers. Regarding the modest claim, the idea is that everyone has inscribed somehow in their mind rules like ‘do not harm <insert subjects here>,’ where what subjects are inserted depends on the culture in which the individual bearing the morality acquisition device develops. Regarding the minimal claim, Prinz suggests the moralization mechanism might convert a tendency to avoid incest into a taboo against incest.

98 Prinz’s other treatments of this topic reach similar conclusions, for example, that “morality...is a by-product of capacities that were not themselves evolved for the acquisition of moral rules” (2007: 270).
disappointingly prosaic: Prinz has fallen foul of the confusion surrounding the concept of innateness.\(^{99}\)

### 3.2.1. Spandrels, Exaptations, and Innateness

That the term ‘innate’ has many meanings should come as no surprise (Griffiths [2002]; [2009]). It is certainly not news to Prinz, who acknowledges the slipperiness of the term and (in a move that would make Pat Bateson\(^{100}\) very happy) says what he means by it:

...a psychological phenotype \(P\) is innate if it is acquired by means of psychological mechanisms that are dedicated to \(P\), as opposed to psychological mechanisms that evolved for some other purpose or for no purpose at all (2008b: 370).

Prinz also recognises that Joyce\(^{101}\) means something different by ‘innate,’ namely, “can be given an adaptive explanation in genetic terms” (Joyce [2006: 2]). The somewhat baffling thing is that, having been careful to specify what is meant by ‘innate’ in the mouths of various interlocutors, Prinz goes on to make the very mistake against which such caution is meant to ward.

That the capacity to make moral judgements is not innate in the sense of not being subserved by dedicated machinery follows, Prinz thinks, from his rejection of immodest, modest, and minimal moral nativism. His problematic move is to then say, on that basis, that the capacity to make moral judgements is not innate in the sense of not being an adaptation. Prinz’s mistake lies (so far as I can tell) in his supposition that being a by-product, being a spandrel, and lacking dedicated machinery are equivalent (cf. the quote above from Prinz [2009: 168]). Careful consideration will show that this is not so.

---

\(^{99}\) A similar charge of sloppiness with the notions of nativism and innateness is laid against Prinz by Neil Levy (2009: 524).

\(^{100}\) Ethologist Patrick Bateson exhorted researchers to “say what you mean (even if it uses a bit more space) rather than unintentionally confuse your reader by employing a word such as innate that carries so many different connotations” (1991: 22). To ‘your reader,’ I would add ‘yourself,’ for reasons that should become apparent as I discuss Prinz’s work.

\(^{101}\) To whom he offers the by-product explanation as part of a critical response. Much the same could be said of Ruse, and others who take the tendency to make moral judgments as the explanans in evolutionary explanations of morality.
To set up the following discussion, a distinction must be drawn between the *capacity* to make moral judgements and the *tendency* to make moral judgements. Prinz runs the two together and, although pulling them apart may seem like hair-splitting, doing so is crucial if we are to see how Prinz’s case for the by-product explanation fails. I will take ‘the capacity to make moral judgements’ to refer to the ability to make moral judgements and to whatever mental apparatus or architecture underlies that ability. I will take ‘the tendency to make moral judgements’ to refer to the practice of actually using the capacity to make moral judgements, and will consider this to be synonymous with ‘the trait of making moral judgements’ and with ‘moralising.’

Having distinguished that capacity for moral judgement from moralising, the next step is to get clear on the notion of a spandrel. Talk of spandrels was introduced into biology by Gould & Lewontin (1979). The term is drawn from architecture, referring to the triangular spaces between the supporting arches of a dome. These spaces are present in buildings not by architectural design but because designed features of buildings necessitate their presence: spandrels “are necessary architectural by-products of mounting a dome on rounded arches” (1979: 581). Gould & Lewontin used ‘spandrel’ to refer to features of an organism that are present, not as the result of natural ‘design’ (i.e. selection for those features) but as a necessary consequence of other features being present. The human chin, for example, must exist, given that the human jawbone develops in the way it does. Chins are spandrels.

Prinz follows Gould & Lewontin in his definition of spandrels: “capacities that emerge as inevitable by-products of other capacities” (2009: 183). Recall, Prinz’s view is that “morality is a spandrel” (2009: 168). If there is no mental machinery dedicated to morality, then it may well follow that the capacity to make moral judgements is a spandrel. Prinz seems to want to say more than that the capacity to make moral judgements is a spandrel, though. He seems to want to say that moralising is a spandrel. For instance, he claims that “the fact that our lives are thoroughly permeated by [moral] norms may be
an accident” (2009: 168). This claim does not follow from the fact that there is no mental machinery dedicated to morality.

Consider: what evidence is there that the making of moral judgements is inevitable – must occur – given that we have certain other capacities? Darwin himself at times suggested something like a spandrel view of morality. He wrote:

Any animal whatever, endowed with well-marked social instincts, the parental and filial affections being here included, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well developed, or nearly as well developed, as in man (2004/1871: 71-72).

Is it really inevitable that social creatures, upon attaining some requisite level of intelligence, will begin to moralise? I hope Darwin would forgive me for doubting it. In any case, there is little in the way of evidence for this necessity claim. We humans have no other comparably intelligent, social creatures here on Earth to examine, nor have any alien species suitable for testing this hypothesis yet deigned to visit us. Of course, Prinz thinks it is not intelligence but, centrally, emotion, out of which moralising springs. So, is it inevitable that emotional creatures will moralise? As it turns out, on Prinz’s favoured account of what moralising is, it really may be inevitable that emotional creatures moralise. Prinz claims that:

...a person regards something as morally wrong (impermissible) if, on careful consideration, she would feel emotions of disapproval [i.e. guilt, shame, disappointment, anger, resentment, indignation, contempt or disgust] toward those who did the thing in question (2009: 179).

Any creatures able to feel emotions of disapproval and disposed to do so with regard to at least some things are bound to moralise, if this is all there is to moralising. The claim that morality is a spandrel thus turns out to depend, I think, on accepting Prinz’s sentimentalist view of what moralising consists

102 Darwin goes on to distinguish between a “sense of right and wrong” and the actions to which that sense attaches, speculating that sufficiently intelligent animals comprising societies very different from our own would (like us) think in terms of what ought and ought not be done but (unlike us) might think that unmarried females ought to kill their brothers, and mothers their fertile daughters (Darwin was imagining intelligent bees).
Here is not the place to debate that view, but I am happy if it has become clear that the spandrel claim is far from established by the empirical evidence Prinz mentions in his discussion of moral nativism.

Even if Prinz is right that there is no dedicated moral mental machinery, that moral judgements are made using capacities evolved for other purposes, the right conclusion to draw is not necessarily that the trait of making moral judgements is a spandrel. The concept to reach for instead may be *exaptation*. Gould & Lewontin employed this concept in their 1979 paper, calling it “secondary adaptation” or the “fruitful use of available parts” (1979: 596, 584). Gould coined the term ‘exaptation’ in a paper written with Elisabeth Vrba, defining exaptations as:

> [U]seful...characters, evolved for other usages (or for no function at all), and later “coopted” for their current role (1982: 6).

Cooptation is a matter of selection favouring a new use of an extant feature. Via this kind of process, adaptations for one purpose, or even spandrels, can be converted into adaptations for a new purpose. For example, black herons (*Egretta ardesiaca*) use their wings not only for flight but also to shade the water in which they wade while hunting fish. This so-called ‘canopy feeding’ reduces glare off the water and makes prey easier to spot. The heron’s wings are adaptations for flight, which were at one stage coopted for use as water-shaders and so became exaptations for hunting and, insofar as...

---

103 Although it is not entirely clear to me, I suppose ‘exaptation’ could also be applied to the case of intelligent agents consciously choosing to employ structures that have one (or no) function for a different and (for the agent) useful purpose. For example, Mack truck drivers talk of ‘following the dog,’ that is, using their vehicle’s characteristic hood ornament as a guide when manoeuvring (apparently, the alignment of the dog with respect to a given reference point gives the driver an accurate overall sense of how his truck is positioned with respect to that point). The function of the dog is, at least as far as Mack truck designers are concerned, purely decorative, but the fruitful use of an available part involved in ‘following the dog’ might sensibly be seen as a case of cooptation of the ornament by Mack truck drivers, and the ornament itself might sensibly be seen as an exaptation (for, I guess, manoeuvring). In any case, I will focus here on exaptations produced by natural selection rather than conscious intention.

104 Gould & Vrba mention this example (1982: 7-8). See also Kushlan & Hancock (2005: 29-31) for a detailed discussion of the various ways in which herons employ their wings while feeding.
selection has since favoured that usage, the heron's wings have become adaptations for hunting as well.

It is perhaps now clear how the notion of exaptation might figure in a response to Prinz's case for the by-product explanation. Allow Prinz the claim that there is no mental machinery dedicated to making moral judgements. Moral judgements would then sensibly be called by-products of whichever capacities are supposed to allow, in conjunction, their production. That would not, however, mean that our tendency to make moral judgements – to moralise – is an evolutionary accident (as Prinz maintains) rather than an adaptation (as is claimed by the likes of Joyce and Ruse). Prinz's supposition that a lack of dedicated machinery means an absence of adaptation is mistaken. As the example of black herons and canopy feeding illustrates, once we avail ourselves of the ideas of exaptation and cooptation, we can make good sense of there being adaptations that are not subserved by dedicated machinery. If we use certain faculties in combination to produce moral judgements and if (as Joyce and Ruse claim) judging in that way has been favoured by selection because of the efficacy of moralising in motivating fitness-enhancing behaviour, then the trait of making moral judgements may be an adaptation even if there is no faculty dedicated to producing moral judgements.105

To recap, Prinz claims that human morality is not "an evolved adaptation" but a "by-product" of capacities that evolved for other purposes ([2007: 263]; [2009: 168]). I have distinguished the capacity to make moral judgements from the tendency to actually do so. I have not tried to settle the complex empirical question of whether humans have mental machinery dedicated to the making of moral judgements. I have allowed for the sake of

105 One may wonder at this point what has become of Prinz's claim that there are no moral universals. As mentioned above (see §3.2), Prinz denies even the minimal claim that humans have a 'moralization mechanism' (i.e. something in our mental architecture that allows us to create distinctively moral rules, the content of which can vary open-endedly across cultures). Above, I granted for argument's sake that there are no moral universals but I am, in fact, inclined to think this minimal version of moral nativism can be defended against Prinz's attack. Prinz's attack on minimal moral nativism relies heavily (by his own admission) on the attack against the moral/conventional distinction by Kelly et al. (2007). I have already responded to Kelly et al. at some length (in §2.3.1). So, I will not go into further detail here regarding moral universals.
argument Prinz’s claim that there is no such machinery. I have tried to show that even if this claim is true, it does not follow that human morality is an evolutionary accident rather than an adaptation, if what is meant by ‘human morality’ is ‘moralising,’ i.e. the making of moral judgements. Moralising, I have tried to suggest, is neither inevitable (hence not a spandrel) nor functionless (hence not a mere accidental by-product).106

3.3. Natural versus Sexual Selection

My discussion thus far has focused on particular issues with SSM. Must we appeal to sexual selection to explain helpfulness toward unrelated unlikely reciprocators, or might kin selection and reciprocity along with the mistaken identity idea suffice? Is it necessary to seek some adaptive rationale for human morality, or could it simply be an accident of evolutionary history that we are creatures who judge in moral terms? My goal thus far has been to situate SSM in relation to other kinds of explanations for helpfulness and for morality. In this section, I will consider the question of how to situate sexual selection in relation to other evolutionary processes.

As mentioned in Chapter 1 (see 1.1.2b), Darwin saw natural selection and sexual selection as distinct evolutionary processes. In what follows, I will begin by laying out Darwin’s distinctions and his reasons for drawing them. Also as mentioned in Chapter 1, Miller follows Darwin in holding natural and sexual selection to be distinct: “evolution is driven not just by natural selection for survival, but by an equally important process... sexual selection through mate choice.” (2000: 3).107

106 Prinz himself says things that suggest he thinks moralizing does have a function, namely, a culturally-selected, group cohesion-promoting function. Perhaps he would not disagree with the line I have taken here. Or perhaps he thinks the products of cultural evolution are not adaptations. In any case, I will not speculate further.  
107 The idea that sexual selection is a distinct and special kind of evolutionary process is maintained in Miller’s later work on SSM. Miller (2007: 98) claims that sexual selection is an especially potent kind of selective force that can “supercharge” other evolutionary processes.
Miller is on one side of a disagreement about how natural and sexual selection stand in relation to each other. On the other side are those who think Darwin’s distinction should be collapsed (with sexual selection being subsumed under natural selection). In this section, I will discuss whether Darwin’s distinction can be maintained, and consider the consequences for SSM if it cannot.

3.3.1. Darwin’s Distinctions

Darwin sought to explain why it is that organisms are so often impressively suited to their habitats and lifestyles. He also sought to explain why organisms sometimes bear features that seem merely ornamental or even encumbering, and that are present in one sex of a species but absent in the other. Like ornaments, some kinds of armaments – horns, spurs, and the like – are also sex-limited. Darwin’s answer to the question of how organisms came to be adapted to their environment was natural selection. His answer to the puzzle posed by sex-limited ornaments and armaments was sexual selection.

Darwin distinguished sexual selection from natural selection on the grounds that:

...[sexual] selection depends, not on a struggle for existence in relation to other organic beings or to external conditions [as does natural selection], but on a struggle between the individuals of one sex, generally the males, for the possession of the other sex. The result is not death to the unsuccessful competitor, but few or no offspring (1998/1859: 117).

According to Darwin, sexual selection “depends on the advantage which certain individuals have over others of the same sex and species solely in respect of reproduction” (2004/1871: 243). Darwin thought there were two ways an individual might gain an advantage in the struggle for reproduction:

The sexual struggle is of two kinds; in the one it is between the individuals of the same sex, generally the male sex, in order to drive away or kill their rivals, the females remain passive; whilst in the other, the struggle is likewise between the individuals of the same sex, in order to excite or charm those of the opposite sex, generally the females, which no longer remain passive but select the more agreeable partners (2004/1871: 398).
Sexual selection is now typically divided into intra-sexual selection (involving the first kind of struggle identified by Darwin in the quote above) and inter-sexual selection (involving the second kind of struggle). As Helena Cronin points out ([1991: 234-235]; see also Andersson & Iwasa [1996: 53]), the terminology here is potentially quite misleading. Intra-sexual selection involves competition between members of the same sex, but so too does inter-sexual selection: the latter does not involve competition between members of different sexes for access to mates (which would, on reflection, be rather strange!). Just where the intra- versus inter-sexual selection distinction comes from is not clear, but a couple of examples will help to illustrate the important difference the terminology is supposed to mark.\(^{108}\)

Elephant seals (*Mirounga leonina*) provide a good example of intra-sexual selection (Darwin [2004/1871]; Andersson [1994]; Fabiani *et al.* [2004]). Male elephant seals fight for breeding territory, with dominant males holding large territories that accommodate many females and thus mating much more than subordinate males.\(^{109}\) Male elephant seals are much bigger than females (indeed, elephant seals are among the most sexually dimorphic of all mammals), and bigger males tend to win territorial disputes with smaller males. The great size of males is thought to be the result of many generations

\(^{108}\) Julian Huxley might be its unwitting progenitor. In a 1938 paper on sexual selection, Huxley talked a lot about 'intra-specific' versus 'inter-specific' competition. Huxley himself never used the term 'inter-sexual selection.' He instead contrasted "intra-sexual selection" with "epigamic selection" (1938a: 431). (Incidentally, Huxley downplayed the role of female choice in the latter, seeing male displays as more the result of selection for male ability to excite passive, coy, reluctant females than of females actively exercising preferences for some males over others.) Even though Huxley didn't refer to 'inter-sexual selection,' it may be that the intra/inter contrast enjoyed a memetic advantage over the intra/epigamic one.

\(^{109}\) Intra-sexual competition is not confined to making threats and fighting for access to mates. It can occur even after many males have mated with the same partner. In bonobos, for example, intra-sexual selection occurs mostly in the form of sperm competition (Marvan *et al.* [2006]). Sperm competition has produced baroque and somewhat eye-watering features in insects. The penis of the male dragonfly has elaborate spines for removing sperm that a female may have stored from her previous mates (Cordoba-Aguilar *et al.* [2003]). The seminal fluid of male fruit flies contains proteins that both incapacitate the sperm of other males and reduce females' inclination to mate again (Chapman *et al.* [1995]). These kinds of morphological and physiological features would count as products of intra-sexual selection as Darwin intended the term.
of bigger males out-reproducing smaller males in virtue of holding larger harems. Male size, that is, is the product of intra-sexual selection.¹¹

Peacocks (*Pavo christatus*) serve as the classic example of inter-sexual selection. Darwin explained the evolution of the male peacock's elaborate train by appealing to the sexual preferences of peahens:

> Just as man can give beauty, according to his standard of taste, to his male poultry... so it appears that female birds in a state of nature, have by a long selection of the more attractive males, added to their beauty or other attractive qualities (2004/1871: 246).

Peahens do in fact prefer to mate with more highly ornamented males (Petrie & Halliday [1994]). Inter-sexual selection is taken to explain a diverse array of otherwise puzzling phenomena, including not just morphological extravagances like the peacock's tail but bizarre behaviours as well, such as the male satin bowerbird's (*Ptilonorhynchus violaceus*) obsessive collecting of blue things with which to decorate his nest (Coleman *et al.* [2004]).

Darwin's presentation of sexual selection left two questions unanswered. Why is it typically males that fight or display to obtain access to females, rather than females fighting or displaying to obtain access to males? Why would the choosy sex in inter-sexual selection (typically the female) prefer such seemingly arbitrary things as flashy tails or an abundance of blue stuff?

¹¹ Darwin recognised that sexual dimorphisms are sometimes the result of natural selection operating differently on males and females. For example, he mentioned huia (*Heteralocha acutirostris*), a now-extinct species of New Zealand wattlebird ([2004/1871: 243]; see also <http://www.nzbirds.com/birds/huia.html>). Huias ate insects and foraged in male/female pairs. Males had a short, stout, strong beak good for breaking apart rotting wood to expose insects. Females had a long, slender, curved beak good for picking insects out of hard-to-reach places. By specializing into different niches, male and female huias competed less for food. Darwin allowed that sexual dimorphism resulting from niche differentiation was explicable just in terms of natural selection. However, there is no evidence that niche differentiation is responsible for sexual dimorphism in elephant seals. Interestingly, though, there is some evidence that in grey seals (*Halichoerus grypus*), another sexually size-dimorphic species, males and females pursue very different foraging strategies, with larger males exploiting regions and resources unavailable to smaller females (Breed *et al.* [2006]). Of course, the fact that males and females have different foraging habits may well be the result, not the cause, of sexual size dimorphism.
The answer to the first question, at least in broad outline, was provided by the work of Trivers (1972) and of A.J. Bateman (1948). Trivers explained female sexual choosiness as the result of the females of a species typically investing more in their offspring than do the males. Bateman noted that males more so than females benefit from mating multiple times (since multiple matings mean multiple offspring for males but not for females) and he thought this accounted for the prevalence of male-male conflict over access to mates.

Several kinds of answer to the second question have been proposed. According to the so-called the ‘good genes’ view (Hamilton & Zuk [1982]; Kirkpatrick & Ryan [1991]), female preferences for elaborate tails or extensive blue-thing collections are not, despite appearances, arbitrary. Rather, pickiness pays, since the impressiveness of a male’s display positively correlates with his genetic quality. The ‘good genes’ label is applied to cover cases in which the benefit to females of choosing males with impressive ornaments or displays is ‘indirect,’ coming in the form of genes contributed by the male to his offspring. When the choosy female gets ‘direct’ benefits, such as the food and protection that a high-quality male can effectively provide, the term ‘good parent’ sexual selection is used (Iwasa & Pomiankowski [1999]).

Standing in contrast to ‘good genes’ and ‘good parent’ views are views on which the male traits preferred by females have no value other than their attractiveness to females. Most prominent among these is the “runaway” selection explanation (Fisher [1930: 136-138]; Kirkpatrick [1987: 58]). According to this explanation, female preferences can drive the spread and exaggeration of male traits via a positive feedback loop in which preferences

---

111 Ronald Fisher (1930) had earlier discussed parental investment in offspring, but in the context of explaining why the sex ratio at birth in most species is approximately 1:1, rather than of sexual selection.

112 Trivers defined ‘parental investment’ as “any investment by the parent in an individual offspring that increases the offspring’s chance of surviving (and hence reproductive success) at the cost of the parent’s ability to invest in other offspring” (1972: 139). Notice, parental investment is not merely a matter of gamete size (i.e. eggs versus sperm) but also includes feeding and care of offspring, so assessing relative male and female parental investment is not straightforward.
and traits become genetically linked. Like the runaway account, "sensory exploitation" accounts do not suppose that flashy male traits indicate good genes or good parenting abilities. Rather, these accounts claim, females prefer certain traits in males simply because those traits appeal to pre-existing biases in the females' sensory systems (Ryan [1990: 179]).

The 'good genes' account of mate preferences has received substantial empirical support. For example, work on peacocks (Petrie [2002]) and on satin bowerbirds (Doucet & Montgomerie [2003]) supports an explanation for female choice in terms of indirect genetic benefits. Accounts of mate choice based on runaway and sensory bias have also been empirically supported. Andreas Paul (2002: 885) reviews evidence for runaway sexual selection in fruit flies, sand flies, crickets and guppies. Alexandra Basolo (1990; 1995) has argued from phylogenetic and experimental data that the long 'swords' of male green swordtail fish (Xiphophorus helleri) evolved to exploit a bias in female swordfish sensory systems.

---

113 For an accessible explanation of the details of runaway selection, see Haufe (2008: 122-123). Runaway sexual selection is sometimes referred to as the 'sexy son' hypothesis. Technically, though, the two ideas are distinct. Weatherhead & Robertson (1979) proposed the 'sexy son' idea to explain why many female red-winged blackbirds (Agelaius phoeniceus) mate with males holding poor-quality breeding territories, or, to pose the puzzle differently, how males with poor-quality territories come to hold large harems (as Weatherhead & Robertson had discovered in their earlier [1977] work). This is a puzzle because females mating with such males take a double hit to their offspring production, since their breeding territory is not only poor but also crowded. Weatherhead & Robertson had found that harem-holding males with poor territories, far more often than males with better territories, engaged in a "song-spread display" toward females (i.e. singing while simultaneously flaunting bright red shoulder feathers). Weatherhead & Robertson thought this display attracted females. Experimental manipulations supported their idea (1977: 589-590). Weatherhead & Robertson (1979: 207) reasoned that a female who mates with a frequently-displaying male may have fewer offspring than a female who breeds in a better territory, but her sons will inherit their father's tendency to display frequently, meaning those sons will eventually attract many females as well. In 'sexy son' sexual selection, then, females lose out in the reproductive short-term (i.e. produce fewer sons and daughters) in order to gain down the generational line (i.e. have more grandchildren), but such 'generation-jumping' is not a feature of Fisherian 'runaway' selection.

114 Möller & Alatalo (1999) review evidence for 'good genes' mate choice in many other species.

115 The details of Basolo's work are fascinating. In essence, she shows that the female swordtails' preference for long-sworded males pre-dates the evolution of male swords by experimentally demonstrating that such a preference is present in females of a closely-related species (Xiphophorus maculatus), the males of which lack swords.
We now have in place Darwin’s distinction between natural and sexual selection, as well as some of the central subsequent work developing Darwin’s ideas about sexual selection.\textsuperscript{116} I must now forestall a potential confusion. Doing so is worthwhile in its own right but will also put in place important groundwork for my discussion of the empirical commitments of SSM in Chapter 5 (see § 5.2.2). The potential confusion concerns just what sexual selection theory claims. My foil here is Joan Roughgarden ([2005a]; [2005b]; [2006]).

Roughgarden casts sexual selection theory as the claim that “passionate males with cheap sperm pursue coy females with expensive eggs” (2005a: 18). She calls sexual selection a “myth” on the grounds that there are just too many cases in which competitive, flamboyant females fight over (and for the attention of) coy, choosy males. As examples of such sex role reversal, Roughgarden lists “mormon crickets, bush crickets, and katydids; the two-spotted goby, and North-Sea pipefish (relatives of seahorses); and among birds, the wattled jacana, red-necked phalarope, and spotted sandpiper” (2005a: 20).

While Roughgarden is right that sex role reversal is far from rare, she is wrong to think this shows sexual selection does not exist. Her mistake is to suppose sexual selection theory is fundamentally a claim about male versus female behaviour. Trivers’ discussion of parental investment is instructive here:

\textit{What governs the operation of sexual selection is the relative parental investment of the sexes in their offspring.} Competition for mates usually characterises the males because males usually invest almost nothing in their offspring. Where male parental investment per offspring is comparable to female investment one would expect...female choice to be no more discriminating than male choice. Where male parental investment strongly exceeds that of the female...one would expect females to compete among themselves for males and for males to be selective about whom they accept as a mate (2006/1972: 141).

\textsuperscript{116} It is perhaps obvious, but I will emphasize at this point that the kind of sexual selection supposed to be at work in SSM is inter-sexual selection of the ‘good genes’ and ‘good parent’ variety. In addition, Miller suggests that conspicuously costly and helpful behaviours might serve as “good partner” displays, that is, reliable indicators of “traits that promote efficient coordination and high mutual benefits in long term sexual relationships” (2007: 101).
For at least some of the species Roughgarden lists, there is evidence that male parental investment exceeds that of females (for example, Gwynne’s [1981] study of mormon crickets). For none of the species listed does Roughgarden give evidence that there is a mismatch between sex role and parental investment, as she would need to do to truly explode the “myth” of sexual selection. Andreas Paul (2002: 882-883) offers a judicious review of the evidence on sex role reversal, relative parental investment, and the benefits of polyandry (i.e. females mating multiple times). The upshot is that there are currently puzzling cases for the theory of sexual selection but not yet any glaring counter-examples.

That Roughgarden’s attack is misdirected is not, I think, entirely her fault. Writers on sexual selection have sometimes made out like the process produces a competitive male/choosy female pattern with only a very few exceptions. Witness Bateman (1948: 365): “in unisexual [i.e. non-hermaphroditic] organisms there is nearly always a combination of an undiscriminating eagerness in the males and a discriminating passivity in the females.” Also, it should be recognised there is definite merit in Roughgarden’s work. She draws attention to many complex and interesting mating systems. She fails, however, to establish that sexual selection (properly understood) does not exist.

3.3.2. Are Natural and Sexual Selection Distinct Processes?

In this section, I will discuss the question of whether sexual selection ought to be recognised as an evolutionary process in its own right, distinct from natural selection. Answering this question requires having some handle on the prior issue of what it is for processes to be distinct. So, what are good grounds for holding two evolutionary processes to be distinct?

117 I wish to set aside an interesting but tangential issue, namely, that of sexual selection’s place in the history of ideas. Helena Cronin (1991) has analysed the historical unpopularity of Darwin’s ideas about sexual selection, especially inter-sexual selection. Sexual selection is no longer a neglected topic – far from it – but this historical story is not the relevant one when I ask whether sexual selection ought to be recognised as an evolutionary process in its own right.
Recall Darwin's claim that sexual selection "depends on the advantage which certain individuals have over others of the same sex and species solely in respect of reproduction" (2004/1871: 243). The "solely in respect of..." locution needs attention. Suppose we replaced 'reproduction' with, say, 'avoiding meteorites.' Suppose also that individuals really did differ in ways that gave some an advantage over others in just this very limited respect. We would not thereby be inclined to suppose that, in addition to natural selection, organisms are subject to meteorite selection.

The general point is that finding a property that some instances of selection have and other instances of selection lack does not suffice to substantiate the claim that there are distinct selective processes at work. Brief consideration shows that if this were sufficient, we would swiftly be faced with an absurd proliferation of kinds of selection, a proliferation limited only by our creative ability to gerrymander distinguishing properties.

The foregoing suggests a need to articulate the debate more carefully. We might do this by marking off talk about processes that 'can be distinguished' from talk about processes being 'distinct.' Let us suppose that processes 'can be distinguished' just in case there is some difference between them. Of course, just about every process can be distinguished in this sense (e.g. making toast versus making toast in France: the latter process can only take place in France, whereas the former can occur elsewhere). But this is precisely the point: Darwin did not just mean to say there is some difference between selection involving interaction between organisms and their potential mates and/or sexual rivals and other instances of selection. That much is evident. That is not denied by those who wish to collapse Darwin's distinction. To say that these two processes are distinct is to say something more than merely that they can be distinguished (in the sense defined above). What more does it take to have a distinction between processes, rather than just to have processes that can be distinguished? Here is my proposal: a

\[118\] I do not think this is likely, of course; the fantastic example is just meant to make my point vivid.
genuine distinction between kinds of evolutionary processes must do worthwhile theoretical or empirical work.

At this point, it would be good to have an uncontroversial example of two distinct evolutionary processes. Drift and selection are, I think, paradigmatically distinct evolutionary processes. Changes in populations over generational time are sometimes random, without overall pattern. Drift explains such changes. Changes in populations over generational time are sometimes non-random. The overall pattern exhibited can be directional, stabilising, or separating. Selection explains non-random change in populations over generational time. Drift and selection are uncontroversially distinct processes, I think, because each is needed to account for a phenomenon the other cannot explain.

Darwin’s original motive for holding natural and sexual selection to be distinct was a good one, by the criterion just laid out. There were, Darwin thought, two phenomena that must be explained: the fit of organisms to their environments, and the existence of sex-limited ornaments and armaments. To do this explanatory work, he thought, it was necessary to posit two processes.

We are now in a position to see why some biologists have been inclined to collapse Darwin’s distinction. That some biologists have been so inclined has been the case since Darwin proposed his distinction right down to the present day. Alfred Russell Wallace attempted to dispense with sexual selection entirely, in favour of purely natural-selection based explanations (Cronin [1991:123-136]). Less extreme but still hostile to the idea was Julian Huxley, who thought sexual selection was “merely an aspect of natural selection which...is concerned with characters which subserve mating, and are usually sex-limited” (1938b: 34). Fast-forwarding, we find Arnqvist & Rowe

---

119 In the first case, individuals close to one extreme with respect to a given trait come to predominate over individuals further from that extreme. In the second case, individuals that are intermediate with respect to a given trait come to predominate over individuals close to either extreme. In the third case, individuals close to either extreme with respect to a given trait come to predominate over average individuals.
claiming that the division between natural and sexual selection is largely historical, and that:

...there is no obvious reason to see sexual selection as truly distinct from natural selection [and] many or most evolutionary biologists today view sexual selection as a component of natural selection.

Marc Hauser (1998) is one such biologist, who clearly expresses the view that sexual selection is not an evolutionary process distinct from natural selection:

[Although there is an important historical distinction to be drawn between natural selection and sexual selection...once you break the problem down into one focused on characters leading to fitness advantages (i.e., gene replication), the distinction really fades away.

These quotes are merely representative; a brief trawl through the literature would easily net more. The Hauser quote is particularly revealing in that it gives insight into why one might deny that natural selection and sexual selection are distinct processes. Recall, Darwin distinguished the two because he took each to be necessary to account for a different phenomenon (adaptedness and sex-limited ornaments and armaments, respectively). Darwin drew his distinction, though, knowing nothing of genetics.

Ernst Mayr attributes the collapse of Darwin’s distinction to the advent of the gene’s eye view of evolution, and of mathematical population genetics in particular:

Sexual selection was entirely rejected by the mathematical population geneticists, who considered evolution as a change in gene frequencies and defined fitness simply as the contribution of a gene to the gene pool of the next generation. Since this definition, indeed, applies equally to natural and sexual selection, any distinction between the two kinds of selection is obliterated (1985: 596).

The key idea here is rather appealing. It is that one process, namely, the differential replication of genes, can do the explanatory work Darwin thought required both natural selection and sexual selection.

How successful an organism is in propagating its genes depends on several factors, standardly called “components of fitness” and labelled ‘viability,’ ‘mating success,’ and ‘fecundity’ (Hamilton [2009: 190]). The first label refers to an organism’s prospects of surviving until reproductive age, the second to the number of mates an organism obtains, and the third to the number of offspring an organism produces per mating. For example, a slightly more efficient metabolism may increase a peacock’s viability, while more lavish tail feathers may increase his mating success. Any increase in an organism’s viability, fecundity, or mating success will \((\text{ceterus paribus})\) increase its fitness. If peahens prefer lavishly-tailed males, then such males will be fitter and may come to predominate over more stingily ornamented ones. If larger elephant seals tend to defeat smaller ones in territorial battles and mate more as a result, then male body size may increase over generational time.

The foregoing suggests the following argument for collapsing Darwin’s distinction. If natural selection and sexual selection are supposed to be distinct on the grounds that each is necessary to explain a different class of phenomena, then showing that all the relevant phenomena can be explained by one kind of process removes the grounds for thinking natural selection and sexual selection are distinct. Differential replication of genes (in virtue of differences between organisms – the genes’ vehicles – in the components of fitness) can explain what Darwin thought it took both natural selection and sexual selection to explain. So, the distinction between natural selection and sexual selection should be rejected.

On this kind of picture, ‘natural selection’ refers to the process of differential replication of genes. Darwin’s ‘sexual selection’ just picks out causes of variance in mating success, rather than some further, distinct evolutionary process. Elliott Sober’s (1993b: 50) distinction between “source
laws” and “consequence laws” in biology is perhaps useful here in understanding the role of sexual selection. Consequence laws in biology describe what happens to a population, given that the fitnesses of its members are as they are. As Sober notes:

[The consequence laws concerning natural selection are preeminently part of the province of population genetics. It doesn’t matter to the equations in population genetics why a given population is characterised by a set of selection coefficients, mutation and migration rates, and so on. These values may just as well have dropped out of the sky. The equations of population genetics merely take these numbers and compute changes in gene frequencies (1993b: 59).

Source laws in biology describe the circumstances that generate fitness differences between members of a population. The various considerations raised by sexual selection theory are, in Sober’s terms, source laws: they tell us an important part of the story about why individuals vary in fitness. Sexual selection is not another kind of consequence law (in addition to natural selection) on this picture, but neither are mate choice, male-male competition and the like unimportant.

Distinguishing between components of fitness was an important move in the case just made for collapsing Darwin’s distinction. Distinguishing between components of fitness also helps undermine another kind of reason that might be offered for thinking natural and sexual selection must be distinct. I will discuss this reason now, since it crops up too often in discussions of natural and sexual selection to simply ignore. I have in mind here the prevalence of talk of conflict between natural and sexual selection.

Arnold (1983: 70) maintains that “the main reason for [considering] sexual selection [to be] distinct from natural selection is that...sexual selection and natural selection can be opposing processes.” Kirkpatrick (1987: 46) in a review of work on sexual selection claims that “a number of field studies...have directly demonstrated cases in which natural and sexual selection are opposed.” Following up Kirkpatrick’s references, we find biologists saying such things as:
[Like female tungara frogs, frog-eating bats are also attracted preferentially to male tungara frog calls with chucks and this phenomenon needs to be interpreted in the context of two opposing selective forces (Ryan [1985:165]).

[Male guppy] colour patterns... represent a balance between selection for crypsis by predators and selection for conspicuousness by sexual selection (Endler [1980: 77]).

Such talk seems to presuppose that sexual selection and natural selection are distinct. After all, if sexual selection were not supposed to be distinct from natural selection, then talk of opposition or balance between the two would be nonsensical, amounting to the claim that a single process can conflict with itself.

In trying to understand the significance of biologists' talk about conflict between natural and sexual selection, we must get clear on just what goes on in purported cases of conflict. If so-called 'conflict' or 'opposition' between natural and sexual selection can be understood without having to suppose there are distinct evolutionary processes operative, then conflict-talk will turn out to be no grounds for maintaining Darwin's distinction.

I think that conflict talk can be cashed out in a way that does not require positing distinct processes at work. One way to do it is to distinguish between components of fitness. In the case of the frogs and fish mentioned in the quotes above, increasing mating success (by making chuck-calls or by bearing conspicuous coloration) decreases viability (by making death by predation more likely). There is no great mystery here, and certainly nothing that requires positing extra evolutionary forces. Conflict talk thus turns out, I think, to offer no support for Darwin's distinction between natural and sexual selection.

Drawing a parallel between sexual selection and kin selection may help to make the point here. Consider again the fig wasps mentioned above (see §3.1.1a). West et al. (2001) say that "competition between relatives can counteract kin selection for altruism" despite the fact that "kin selection favours altruism [when] competing males are more highly related" (2001:...
We needn’t posit two processes (e.g. kin selection and individual selection) to make sense of this conflict talk. We need only observe that, in certain situations, the trade-off involved in benefiting one’s relatives at a cost to oneself is simply not worth making.\textsuperscript{121} Recall Hamilton’s Rule: helpfulness may evolve if $rb > c$ (see footnote 91 in §3.1.1a). What West \textit{et al.} show is that, for a given value of $b$, increasing $r$ will not necessarily lead to increased helpfulness because $c$ may also increase. Kin selection is not being ‘opposed’ or ‘counter-acted’ by some other selective force. Rather, it just turns out (in the fig wasp case) that limited dispersal increases $r$ but also increases $c$, thus having a mixed effect on the inclusive fitness calculation by increasing both sides of the inequality.

At this point, one determined to maintain that natural selection and sexual selection are genuinely distinct processes might claim there really is some phenomenon that requires positing sexual selection as an explanation. Funnily enough, Richard Dawkins makes just such a claim.\textsuperscript{122} He writes:

\begin{quote}
What sexual selection explains, better than natural selection, is diversity that seems arbitrary, even driven by aesthetic whim. Especially if the variation concerned is geographical. And also especially if some of the features concerned ... differ between the sexes (2004: 77).
\end{quote}

Mary Jane West-Eberhard (1983) has made a similar suggestion, that sexual selection is crucial in explaining taxonomic patterns, in particular, very rapid speciation in certain lineages. In the same vein, Price (1999: 101-102) collects together empirical work showing that bird groups “which are

\textsuperscript{121} Recognizing the ubiquity of trade-offs is another way (in addition to distinguishing between different components of fitness) to cash out talk of conflict. Even focusing on just one component of fitness – viability, say – it will often be the case that competing demands pull an organism in different directions. Leaf shape in plants is a good example here. Leaves gather sunlight for photosynthesis but also lose water to evaporation. The optimal leaf shape for gathering sunlight is broad and flat, while the shape that minimizes water loss is spiky, and actual leaf shapes reflect a trade-off between these two considerations. This does not, however, mean that plants are subject to both light selection and water selection. (Thanks to Kim Sterelny for this example).

\textsuperscript{122} I say ‘funnily enough’ because Dawkins tends to subsume sexual under natural selection, writing, for example, that “Dennett has credited Darwin with the greatest idea ever to occur to a human mind. This was natural selection, of course, and I would include sexual selection as part of the same idea” (Dawkins [2004: 66]).
apparently subject to strong sexual selection are particularly speciose.” It is worth taking a moment to provide some setting for these ideas.

Consider: organisms that are found on widely separated branches of the phylogenetic tree but which occupy similar niches can display striking morphological similarities. For example, take the convergences between bat and pterodactyl wings, shark and dolphin fins, and the snouts of the armadillo, the pangolin, and the anteater.

Consider now the bellbirds of South and Central America. There are four species of bellbird in the genus *Procnias*: the white bellbird, the three-wattled-bellbird, the bearded bellbird, and the bare-throated bellbird (Kricher [1999: 272-273]). All four species are rainforest-dwelling frugivores but, as their names suggest, there are vivid differences in morphology across the genus.123

Consider also the spectacularly diverse cichlid populations of Africa’s Great Lakes. At least 600 species of cichlids have arisen in Lake Malawi, which is only one or two million years old, while the even younger Lake Victoria (at a mere 12,000 years of age) has spawned more than 500 cichlid species. Many of these species are genetically compatible but reproductively isolated in virtue of different coloration and mating displays (Johnson *et al.* [1996]; Goldschmidt [1998]; Turner [2007]).124

---

123 At least when it comes to the males; bellbird females are uniformly drab. Consider also, the exuberant morphological diversity on display within the bellbird genus recurs at the next highest taxonomic level. Bellbirds belong to the family Cotingidae, which also includes umbrella birds (genus *Cephalopterus*) and cocks-of-the-rock (genus *Rupicola*). Umbrella birds and cocks-of-the-rock are, like bellbirds, rainforest-dwelling frugivores. Male umbrella birds have a fan-shaped plume of black head feathers and an inflatable air sac dangling over their breast. Cocks-of-the-rock have a head plume that is brilliant orange rather than black, no breast sac, and a fringe of delicately elongated feathers on each wing. Neither umbrella birds nor cocks-of-the-rock have bellbird-like wattles.

124 One of the authors referenced above — Tijs Goldschmidt — recounts his first-hand experience of the overwhelming diversity present in Lake Malawi:

> I myself once caught an unnamed, exceedingly lively, fiery-tempered male with purple flanks and a pitch-black mask but then let him go because at the time I couldn’t face discovering another species. I don’t think it was ever caught again! (1998: 5)
Recall for a moment our paradigm case of distinct evolutionary processes: drift and selection. Drift won’t do to explain directional, stabilising, or separating patterns in populations over time; some kind of selective process is needed. The problem with the proposal that Darwin’s distinction be maintained because sexual selection is needed to explain certain geographical and taxonomic patterns of diversity is that it is just not clear that natural selection (*sensu* Darwin) can’t do the job alone.

Andersson (1994: 210) points out that the diversity on display in groups like bellbirds and cichlids may be the result of adaptive radiation into different (even marginally different) ecological niches followed by the evolution of species-specific mate recognition systems. Species recognition traits do not fit the profile of sexually-selected traits as Darwin understood them. Such traits don’t give an advantage solely in regard to mating; being recognised as a fellow by other members of one’s species may be a good thing for all sorts of reasons. Nor does the development and display of such traits involve any kind of competition between members of the same sex (indeed, the whole point of such traits would be undermined if males started trying to one-up each other and thus produced individually variable forms of the trait).

To recap: finding a property that some instances of selection have and other instances of selection lack does not suffice to substantiate the claim that there are distinct selective processes at work. A genuine distinction between selective processes must do some useful work. Darwin’s distinction between natural selection and sexual selection was well-intentioned (since he thought two distinct processes were required to explain two distinct puzzles) but advances in biology have made the distinction appear unsustainable.

3.3.2b. Utilitarian versus Signal Selection

Biologist Amotz Zahavi was introduced in Chapter 1 (see §1.2.2c) as the architect of the handicap principle. As mentioned, Zahavi’s ideas about signalling were outré when first proposed and were not accepted for some time
afterward. In this section, Zahavi reprises his role as pariah. Will the mainstream once again engulf him or will he instead be left high and dry? In this case, I think the latter is more likely.

Zahavi’s contention, as laid out in Chapter 1 (see §1.2.2c), is that signals must be costly if they are to be reliable. As a result, he claims:

...the fundamental difference between signals and other traits...is the relationship between the signal and its cost. ... In traits other than signals, the cost of the trait is an unavoidable side effect. In signals, cost is of the very essence; it is necessary to the existence of the signal (Zahavi & Zahavi [1997: 59]).

Given this difference between signals and other traits, Zahavi thinks ‘natural selection’ should be used as an umbrella term to cover two fundamentally different selective processes:

[N]atural selection encompasses two different, and often opposing, processes. One kind of selection favours straightforward efficiency, and it works in all areas except signalling. This selection makes features — other than signals — more effective and less costly. We suggest calling it “utilitarian selection”. The other kind of selection, by which signals evolve, results in costly features and traits that look like “waste”. It is precisely this costliness, the signaler’s investment in the signals, that makes signals reliable. We suggest calling this process “signal selection” (Zahavi & Zahavi [1997: 40]).

According to Zahavi’s latest work on the topic, utilitarian selection is a process that “tends to increase the efficiency of traits,” while signal selection “reduces the efficiency of certain traits for the sake of reliability” (2007: 155).

‘Efficiency’ is never explicitly defined by Zahavi, but the way he uses the term makes clear what he has in mind: it is a measure of functional performance relative to cost. To take a simplified example, imagine a mollusc’s shell, suppose its function is to protect against bird predators, and suppose also that a shell’s protectiveness depends on how resistant the shell is to breaking when dropped onto rocks. If shell type A requires a given amount of material \( m \) to construct and breaks 20% of the time when dropped, while

125 The changing tone of Richard Dawkins’ discussions of handicap signalling in successive editions of *The Selfish Gene* provides an amusing index of Zahavi’s biological currency.
shell type $B$ also costs $m$ to build but breaks only 10% of the time (perhaps because it is differently shaped), then $B$-type shells are more efficient than $A$-type shells.

'Reliability' is similarly straightforward: it is a measure of how well signalling correlates with the possession of a given property (e.g. how well the adoption of a threat posture correlates with fighting rather than fleeing when challenged. Slightly more complexly, we might see reliability as a matter of how well signalling at a particular intensity correlates with the possession of a given property to a particular degree (e.g. how well the volume and frequency of a chick's begging calls correlates with the chick's hunger level).

Before critically evaluating Zahavi's distinction between utilitarian selection and signal selection, there are two clarifications that must be made. These concern the distinction itself and the import of the distinction in my overall discussion.

Zahavi's claim that costs are "the very essence" of signals and "necessary for their existence" (1997: 59) requires some unpacking. Even a peacock's tail, the exemplar of a costly signal, can be grown and displayed while imposing absolutely no burden on the peacock. This happens all the time in captive populations, where food is abundant and predators absent. Extravagant tails on captive peacocks seem to falsify Zahavi's claim. Meanwhile, on the other side of the distinction, we may worry that Zahavi's claim about traits other than signals is that such traits would still have evolved even if they were not costly (since cost is not "the essence" of such traits). This would be a worry because evaluating claims of the form 'had trait $T$ involved no cost, $T$ would have evolved anyway' requires us to imagine strange worlds quite remote in modal space, worlds in which, for instance, lactating does not at all deplete females' nutritional resources, or where hunting heavily armed and armoured prey poses no risk at all to predators.

Extravagant tails on captive peacocks would falsify Zahavi's claim if that claim were about signal tokens. The claim is not, however, that every
token of a signal type must inflict costs on the signaller. Rather, it is that a
given type of signal could not persist over evolutionary time were it not
typically costly enough to make it reliable enough to be worth attending to.
And happily, there is no need to strain our intuitions by considering strange
and magical scenarios like those described above. Zahavi’s claim about the
different role played by costs in utilitarian versus signal selection actually
generates clear and (at least in principle) empirically testable predictions:

When the investment required for the development of a particular
utilitarian trait is reduced, the use of that trait may increase, or it may not
change. It is not dependent on the fact that other individuals can use it
equally well. This is not the case with signals. In order to be reliable, signals
require differential investments. If the investment in the reliability of a
signal is lowered to the extent that all individuals can perform it alike - the
signal can no longer provide reliable information on differences between
signallers and will go out of use (2007: 155-156).

Of course, actually testing these predictions is another matter. For
the moment, though, I am content if some confusions about the distinction
Zahavi is trying to defend have been cleared up.

As a final bit of ground-clearing clarification, it must be pointed out
that even if Zahavi is right, Darwin won’t be vindicated. Zahavi’s distinction
between utilitarian selection and signal selection cross-cuts Darwin’s
distinction between natural selection, on the one hand, and intra-sexual
selection and inter-sexual selection on the other. Signal selection covers
signalling of all kinds: alarm calls, begging cries, and warning coloration
(which are naturally-selected sensu Darwin), as well as threat displays
between sexual rivals (Darwin’s intra-sexual selection), and courtship displays
(inter-sexual selection). Utilitarian selection covers everything else: wing
design and cryptic coloration (naturally selected sensu Darwin), weaponry for
use on sexual rivals (Darwin’s intra-sexual selection), and any traits that
function in the context of mate choice broadly construed but which do not

126 Zahavi mentions some fascinating, if at best suggestive, research on satin bowerbirds
(Hunter & Dwyer [1997]). Compared to males that inhabit remote areas, males that live near
human settlements (where blue stuff is far more easily obtained) steal far less from other
males’ collections but engage in far more bower destruction. Stealing from other males is a
way of upping signal intensity, so the change away from that tactic when signalling at a high
intensity is relatively easy looks like support for Zahavi’s claim about what happens to signals
when they start to get cheap.
involve signalling, such as mechanisms allowing post-copulatory sorting of sperm or post-fertilisation selective abortion (Paul [2002: 880]).

Zahavi himself realises this, saying that his distinction “supports Darwin’s intuition that there are two different selection mechanisms that operate in nature [b]ut the differentiation between them is not the same as that proposed by Darwin.” (2007: 144). Still, if Zahavi’s distinction holds up, it could support Miller’s claims about SSM. Recall, Miller claims that SSM is a novel explanation for the evolution of morality because it appeals to a special kind of evolutionary process. If signal selection really is fundamentally different to selection not involving signals, then an account of the evolution of morality that proceeds in terms of signalling really would be importantly different to non-signalling-based accounts. And, interestingly, Miller himself at times contrasts sexual selection with “utilitarian” natural selection (e.g. 2000: 3, 416).

Zahavi’s distinction does not hold up. The first thing to be said against it is that not all signals must be costly in order to be reliable. When sender and receiver interests coincide, there is no need for signals to be costly because there is no incentive for signallers to deceive. This point is nothing new. Zahavi is well aware of this kind of response to his view. His claim in reply is that there is always some conflict of interest between sender and receiver, even between “the most cooperative of parties, such as parents and offspring” (2007: 147). This claim may well be true. Even if it is, though, Zahavi has missed a crucial point. Some interests conflict sometimes, but not all interests conflict all the time. The distinctive distress call given by a lost joey to tell its mother its location is cheap but reliable because both parties share this particular interest, even if other of their interests often (or even always) conflict.

---

127 Although parents and offspring are hardly “the most cooperative of parties,” as has been appreciated at least since Trivers’ (1974) discussion of parent/offspring conflict. It would be interesting to hear Zahavi’s ideas about where the conflict lies in the case of alarm signaling between clonal organisms, as described by Pelletier (2004) in his study of ascidians (colloquially called ‘sea squirts’).
A less often appreciated problem with Zahavi’s distinction is that it breaks down on the utilitarian side as well. Some selection not involving signals, which should be utilitarian if Zahavi is right, in fact does not produce efficiency. I have in mind here what has variously been called a “Red Queen” effect (Van Valen [1973]), “escalation” (Vermeij [1993]), or an evolutionary “arms race” (Dawkins [1979]).

Predators, competitors, or dangerous prey can force organisms to invest more and more in certain traits, and that investment can in turn force those predators (etc.) to invest heavily in counter-measures. The dynamic can be simple, as when trees strive to overtop each other in a race for sunlight. The dynamic can be more complex. For example, mollusc shells may become thicker and spinier to deter hungry crabs, which respond by developing tougher and more powerful claws, which drives molluscs to become thicker and spinier still. The result is that shells and claws get more costly without a corresponding gain in effectiveness (powerfully-clawed crabs nowadays dine no better than did their less formidably armed ancestors in times when shells were thinner). Vermeij even argues that escalation of this kind leaves species absolutely worse off, since the fossil record indicates that “highly escalated” marine gastropod species (i.e. those that have evolved many expensive counter-adaptations to the traits of other organisms) are less likely to survive mass extinction events than less escalated species (1993: 408). Whether Vermeij is right or not, the point against Zahavi is this: some of what he must count as utilitarian selection (because it does not involve signalling) actually fails to produce efficient traits.

Zahavi might offer the following in reply. In evolutionary arms races, utilitarian selection favours the most efficient trait at a given time in the circumstances then obtaining. Over time, efficiency may admittedly decline, but this is because utilitarian selection is not foresighted. At no point is utilitarian selection favouring inefficiency. It favours maximum benefit at minimum cost, and sometimes the best an organism can do is run in place.
I grant that such a reply would be a sensible way of saving Zahavi's claim that utilitarian selection shapes traits to be more efficient and less costly. However, this reply itself does damage to Zahavi's distinction. Signal selection also favours maximum benefit at minimum cost. It's just that, when it comes to signals, the minimum cost may still be quite high, given the need to maintain signal reliability. There seems now to be no principled difference between Zahavi's two kinds of selection.

Summing up: Zahavi's distinction between utilitarian and signal selection held out some hope that the kind of selection appealed to in SSM - inter-sexual selection - might be somehow special. Zahavi's distinction does not carve nature at the joints, though. So, yet another proposal for marking off genuinely distinct kinds of evolutionary processes has failed.

3.3.3. The Consequences of Collapse

I have considered at length the question of whether natural selection and sexual selection are, as Darwin thought and as Miller maintains, distinct evolutionary processes. A large part of the task has been to identify various proposals for distinguishing the two and to evaluate these proposals in some principled way. The upshot, I think, is that Darwin's distinction cannot be maintained.

The conclusion of my discussion is not that sexual selection doesn't exist or is unimportant. Rather, 'sexual selection' is best understood as shorthand for a particular array of factors that affect fitness. In much the same way, talk of 'kin selection' serves, not to pick out a fundamentally different kind of selection, but instead to alert us to the importance of inclusive fitness effects.

Miller insists on Darwin's distinction and makes much of the supposedly special nature of sexual selection. However, SSM is still interestingly different to extant accounts of the evolution of morality even if Darwin's distinction collapses. The parallel with kin selection is again instructive. Hamilton's explanation of Hymenopteran eusociality in terms of
inclusive fitness effects (rather than group selection) was new and a great advance in understanding, even though it did not posit the operation of a kind of selection fundamentally different to natural selection.

It remains to be seen whether SSM will be as resoundingly vindicated and as influential as Hamilton’s theory. Nevertheless, Miller has already succeeded in alerting us to the (hitherto understudied) relevance of moral behaviour to an important component of fitness, namely, mating success.

3.4. Conclusion

My goal in this chapter has been to situate sexual selection in relation to other processes typically taken to be important in the evolution of morality. The critical focus early in the chapter was on Jesse Prinz’s argument that morality is an evolutionary by-product. Were this argument to succeed, it would undermine adaptationist accounts of morality, of which SSM is one. My response to Prinz has been to reveal and resolve confusions in the use of the concept of innateness in his argument. Prinz moves, illegitimately, from a claim about the nature of the mental machinery subserving moral judgement to a claim about the evolutionary history of the trait of making such judgements. Prinz’s argument that morality is an evolutionary by-product fails because the tendency to make moral judgements may be an adaptation even if (as Prinz claims) there is no psychological mechanism dedicated to the making of such judgments.

The critical focus shifted later in this chapter to Miller’s claim that sexual selection is a special kind of evolutionary process distinct from natural selection. In attempting to assess this claim, the immediate stumbling block was that the grounds on which we might decide that in some case we have two distinct evolutionary processes at work are far from clear. My proposal is that, in order to be legitimate, drawing a distinction between kinds of evolutionary processes must do worthwhile theoretical or empirical work. A paradigm case of distinct evolutionary processes, namely, selection versus drift, satisfies this requirement.
Having laid out grounds for deciding when two processes are distinct, I then presented numerous arguments, drawn from Miller’s work and the wider biological literature, for the distinctness of natural and sexual selection. I found those arguments wanting. I have concluded that there is probably not a sustainable distinction to be drawn between natural and sexual selection. I have also shown, though, that this conclusion need not vitiate SSM. Even if sexual and natural selection are not distinct evolutionary processes, SSM has a distinctive contribution to make to discussions of the evolution of morality, in much the same way as the theory of kin selection, when first proposed, greatly advanced our understanding of cooperation despite not positing any evolutionary processes aside from natural selection.

This chapter has served to connect my revised version of SSM up with extant work on the evolution of morality in a more detailed and considered way than Miller managed with his version of the view. The overall goal has been to dispel any lingering whiff of greedy theorising. The air, I take it, is now clear.
Chapter 4. SSM: Conceptual Clarification

4.1. Introduction

This chapter continues the project, begun in Chapter 1, of clarifying SSM. I have already drawn several distinctions that are neglected in Miller’s initial (2000) presentation of SSM (see §1.2.2c). In this chapter, I will also distinguish between several different versions of costly signalling theory. As mentioned in Chapter 1, Miller frames SSM in terms of Zahavi’s handicap principle. Costly signalling theory has, however, undergone several important developments since Zahavi proposed the handicap principle. I will show how Miller’s reliance on Zahavi’s initial formulation of costly signalling theory, together with his neglect of the distinctions drawn in Chapter 1, generates problems for SSM. Finally, I will show how framing SSM in terms of a more sophisticated understanding of costly signalling theory can address at least some of the problems I have raised.

By way of further introduction to this chapter, here again are the distinctions drawn in Chapter 1. A cue, recall, is any feature of the world that can be used by an organism as a guide to action, while a signal is a feature that evolved in order to affect the actions of others. A stotting gazelle sends predators a signal; a limping gazelle provides predators with a cue. Recall also, that costly signalling must do more than just expend resources; it must reduce some component of the signaller’s fitness and, crucially, that fitness cost must be higher for dishonest than for honest signallers. Recall, finally, that while some signals (handicaps) are kept reliable by costs, other signals (indices) are kept reliable by constraints, meaning that the impossibility of faking rather than the unprofitability of doing so is what keeps signallers honest.

One last bit of introductory scene-setting is needed here, namely, a statement of the costly signalling idea as it features in Miller’s presentations of
SSM. In his initial (2000) presentation of SSM, Miller gives the following gloss on the costly signalling idea:

[Zahavi] suggested that the only reliable way to advertise one's true fitness is to produce a signal that costs a lot of fitness. This explains why sexual ornaments are so often large, extravagant, costly, and complicated. The peacock's tail is not just a cheap, transient advertisement visible only to peahens. It is heavy, encumbering, hard to grow, hard to preen, and highly visible to predators. Peacocks have to drag it around everywhere they go. Unfit peacocks might be able to grow large tails, but they would not be strong enough to carry them while finding food, or fast enough to escape from predators. Only highly fit peacocks can afford very large tails. Therefore, if a female sees a male sporting a very large tail, she can be confident that he has high fitness, and that his good genes could be passed on to her offspring (2000: 123-124).

In his updated (2007) version of SSM, Miller presents the costly signalling idea in virtually the same way:

Costly signalling theory offers a solution to [the] problem of lying: if a signal is so costly that only high health, high status, high condition animals can afford to produce it, the signal can remain evolutionarily reliable... [S]exual ornaments almost always impose high costs on the bearer, guaranteeing their reliability as indicators of condition and fitness (2007: 100).

I will follow Searcy & Nowicki (2005: 10) in calling this the "Zahavi handicap" version of costly signalling theory. The key claim made by the Zahavi handicap version of costly signalling is that the high cost of signalling imposes a test on signallers, a test that only who genuinely possess the quality signalled can pass. There are other incarnations of the costly signalling idea. We will meet them in due course. For now, the goal is just to make clear how costly signalling is supposed to work in Miller's presentation of SSM.

---

128 As mentioned in Chapter 1 (see §1.2.2c), the book by Searcy & Nowicki along with one by Maynard-Smith & Harper (2003) will be important sources for the discussion in this chapter. Both books survey, widely and in detail, theoretical and empirical work on costly signalling.

129 Compare Zahavi: An individual with a well developed sexually selected character is an individual which has survived a test. A female which could discriminate between a male possessing a sexually selected character, from one without it, can discriminate between a male which has passed a test and one which has not been tested. Females which selected males with the most developed characters can be sure that they have selected from among the best genotypes of the male population (1975: 207).
4.2. Signals and Costs in SSM

With the scene now set, I will work carefully through Miller’s presentations of SSM. My discussion will be structured around three questions. First, what exactly are the signals in SSM, according to Miller? Second, what features are supposedly being advertised by these signals? Third, are the supposed signals appropriately costly?

4.2.1. Sorting Signals

Miller refers to many things as signals in his presentations of SSM. He refers to virtues as signals. He also refers to gifts and charitable donations as signals. Heroism, by which Miller means braving physical dangers for others’ sake, also gets called a signal. Kindness, sympathy, fairness, honesty, conscientiousness and patience all get referred to as signals at various points throughout Miller’s work. Finally, Miller at times suggests that moral judgements are signals.

The claim that virtues are signals is problematic, given Miller’s take on the nature of virtues. Recall, virtues on Miller’s account are “emergent properties” somehow arising out of the interaction between the personality traits of one individual and the preferences of another (2007: 103). If virtues thus understood were signals, then the receiver’s psychology would somehow feature in the generation of the signal. I confess to finding this deeply confusing. Accordingly, I will set this claim aside and focus on the other, more comprehensible claims Miller makes about what the signals are in SSM.

The claim that gifts and charitable donations are signals is developed slightly differently in Miller’s initial and updated versions of SSM. In his initial version, Miller treats gifts and donations as signals both of wealth and

---

130 Miller (2007: 100) writes “many human virtues evolved through sexual selection as costly signals.”
133 e.g. Miller (2007: 101)
134 Miller (2000: 332, 340)
of generosity. Miller focuses on the role of gifts and donations in signalling psychological qualities, such as "altruism, kindness, and sympathy," in his updated version of SSM (2007: 105), leaving material resources mostly unmentioned. Below, I will consider both possibilities.

The claim that heroism is a signal is also open to two interpretations. Miller himself suggests that heroism is a signal of "competence" (2007: 98). Miller does not elaborate but his understanding of heroism as braving physical dangers to save others makes clear what he has in mind. Rescuing someone from a burning building, a freezing pond, or a hungry beast all seem to demand great strength, agility and fitness. Heroism might thus be taken to signal physical quality. Alternatively, heroism might be taken to signal psychological qualities. This second interpretation seems more plausible the more broadly we construe 'heroism.' Standing up to a bully is a form of heroism available to fit and fat folks alike. The former might take the opportunity to show off their physical gifts. The latter might take a beating that makes their physical failings painfully apparent but succeed nonetheless in displaying attractive psychological qualities like courage and a sense of fairness. Our notion of heroic behaviour covers cases that involve no arduous physical test at all: Rosa Parks' refusal to give up her bus seat stands out as a prime example of such heroism. As with gifts and donations, I will consider both possible interpretations of the claim that heroism is a signal.

The claim that kindness, sympathy, fairness, honesty, conscientiousness, and patience are signals badly needs clarification. Miller sometimes writes as if these psychological qualities themselves are signals. For instance, he talks about "moral capacities for conscientiousness and patience" and "moral capacities for empathy and sympathy" being signals (2007: 101). This is problematic. Signals must be perceptible to receivers. Psychological qualities (or capacities) are not perceptible. The claim that kindness (sympathy, etc.) serve as signals must be spelled out in terms of the signalling role of various behaviours. Behaving in paradigmatically kind ways

---

135 In the Charles Atlas rather than the Charles Darwin sense.
signals a certain psychological trait, namely, being kind, while sympathetic behaviour signals a sympathetic nature, and so on for fairness, honesty and patience. This may seem obvious, but getting things straightened out is necessary after Miller has tangled himself up with talk of virtues and capacities serving as signals.

The claim that moral judgements – meaning public moral pronouncements rather than private mental events – serve as signals is not a central part of Miller’s work on SSM. It gets a rather offhanded mention and only a brief discussion in Miller’s initial presentation of SSM. There, Miller writes:

When individuals espouse ideological positions, we typically interpret their beliefs as signs of good or bad moral character (2000: 332).

[M]oral emotions, judgements, and reasoning were favoured during courtship between our ancestors. Their sexual choices were not satisfied with a few tokens of romantic generosity (2000: 340 [emphasis added]).

By “ideological positions” Miller has in mind political and religious views, but the espousal of moral views certainly should not be overlooked as a form of moralistic self-display. While the idea that moral judgements themselves may serve as signals is not followed up in Miller’s later work on SSM, that idea is worthy of careful consideration nevertheless. Publicly praising and blaming, commending and condemning, are common and important ways in which we (intentionally or not) reveal our character to others. I will consider below whether or not this aspect of morality can be fruitfully understood in terms of costly signalling.

At this point, the distinction between signals and cues can be used to state a cleaned-up version of Miller’s claims about the signals and the features signalled in SSM. According to the cleaned-up version of SSM, the signals are various actions: giving goods away, running risks to rescue others, lending a sympathetic ear or a shoulder to cry on, doing at least one’s share of the work and taking no more than one’s share of the profits, telling the truth, and publicly pronouncing on others’ moral fortes and foibles. These actions signal
the material resources and/or physical qualities and/or psychological qualities of the agent. These features in turn serve as cues in mate choice (i.e. features it pays to pay attention to when choosing a mate), because agents with these features make good parents, good partners, or both.

Two of my opening three questions have now been addressed. We now have an account of what exactly are the signals in SSM are supposed to be, and of what features are supposedly being advertised by these signals. The work so far has been straightforward, mostly organisational and clarificatory. Addressing the third question will be more complicated. Recall, that question asks whether the purported signals in SSM are costly in the way that, according to the Zahavi handicap version of costly signalling theory, will ensure reliability.

4.2.2. Counting Costs

Miller (2000: 347) notes that sending signals to potential mates can be costly in terms of “time, energy, risk, and resources.” Miller (2007: 100) says much the same thing: when it comes to maintaining the reliability of a signal, “any fitness-related cost will work: matter, energy, time, or risk.” As already noted, though, there is an important distinction between costs and expenditures. Actions may require the expenditure of time, energy, etc. without imposing fitness costs. Thus, to assess whether the signalling behaviours listed above are appropriately costly, we need a detailed accounting of the expenditures involved as well as some way of connecting those expenditures up to fitness costs.

Searcy & Nowicki (2005) offer a classification scheme for signal costs (by which we should understand ‘expenditures,’ of course). Their scheme provides a useful starting point for thinking carefully about the costs of moral behaviour.

Searcy & Nowicki define receiver dependent (RD) costs as those that “stem from some response of the receiver” (2005: 14). RD costs can come in
the form of the risk that the signaller will be attacked by the receiver (a “receiver retaliation” cost), or, of the risk that the signaller will be injured should the receiver decide to attack (a “vulnerability” cost).

Searcy & Nowicki define receiver independent (RI) costs as those “imposed regardless of whether or how receivers respond” (2005: 15). RI costs are further divided into developmental, production, and maintenance costs. Developmental costs are paid during the time the signal develops. They can include the resources needed to build display structures as well as those needed to build the capacities underlying display behaviours. Production costs are paid when the signal is sent to the receiver. They can include the energy required to send the signal, the time spent sending the signal, and the predation risk that sending the signal entails. Maintenance costs are ongoing costs, the consequences of having to live with a display structure. They can include increased energy expenditure, increased predation risk, and decreased foraging success.

Searcy & Nowicki’s classificatory scheme is presented in Table 1 below:

<table>
<thead>
<tr>
<th>Developmental</th>
<th>Production</th>
<th>Maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receiver-Dependent</td>
<td>Risk of attack</td>
<td>Risk of injury</td>
</tr>
<tr>
<td>Receiver-Independent</td>
<td>Energy Resources</td>
<td>Energy Time Risk of predation</td>
</tr>
</tbody>
</table>

Table 1: Classification of Expenditures (following Searcy & Nowicki [2005])

Searcy & Nowicki’s scheme offers far more guidance than Miller’s brief list when it comes to identifying the expenditures required by moral behaviour. I am not entirely happy with that scheme, though. Consider: the various risks run by a signaller in signalling are distributed on either side of Searcy & Nowicki’s RD/RI divide. Risks posed to the signaller by receivers are RD costs, whereas risks posed to signallers by other observers (e.g. predators) are RI costs. While distinguishing between developmental,
production, and maintenance costs is worthwhile, I am tempted to slightly redraw the picture, as below in Table 2:

<table>
<thead>
<tr>
<th>Developmental</th>
<th>Production</th>
<th>Maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Imposed</strong></td>
<td>Risk of attack (by receiver)</td>
<td>Risk of predation</td>
</tr>
<tr>
<td></td>
<td>Risk of injury (by receiver)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Risk of predation</td>
<td></td>
</tr>
<tr>
<td><strong>Intrinsic</strong></td>
<td>Energy Resources</td>
<td>Energy Foraging difficulty</td>
</tr>
<tr>
<td></td>
<td>Energy Time</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Intrinsic vs. Imposed Expenditures

In this scheme, the major divide is not between expenditures incurred as a result of the receiver’s actions and those not resulting from the receiver’s actions. Rather, it is between expenditures the signaller pays regardless of what receivers or any other observers do, and those that depend on the actions of receivers or other observers. This seems to me a more natural way of grouping expenditures. I have called this the distinction between ‘intrinsic’ and ‘imposed’ expenditures. As a final note, the contents of the cells of the tables above are meant to be illustrative rather than exhaustive of the different kinds of expenditures defined by the row and column headings. Defining the different possible kinds of expenditures is the most important thing.

Having laid out a classification scheme for expenditures, I will now say something in general about connecting expenditures to fitness costs. Making this connection is crucial if a costly signalling claim is to be

---

136 There is an interesting clarificatory point to be made here about which organisms count as receivers of a signal and which are merely observers. Recall Maynard-Smith & Harper’s definition of a signal (see §1.2.2c). According to that definition, the effect of a trait (behavioural or morphological) on others’ behavior had to be part of the explanation for why that trait evolved in order for the trait to count as a signal. This suggests that it is organisms whose responses help explain the adaptive value of a trait to the signaller that should be counted as the receivers of the signal. Other organisms, whose responses to the signal do not help explain why the signaller evolved that particular behaviour or morphological trait, should be counted as mere observers. This seems to me the right way, in principle, to divide receivers from observers. In practice, telling between the two may sometimes be difficult. For example, stalking lions are certainly receivers of a stotting gazelle’s signalling behavior, but what of the gazelle’s conspecifics? Perhaps nearby sexual rivals and potential mates note the gazelle’s athleticism and alter their behaviour toward it (e.g. yielding to it in confrontations and accepting its advances, respectively), thus providing additional selective pressure in favor of stotting and qualifying themselves as signal receivers.
substantiated. However, direct evidence of fitness costs is very rarely provided in work on costly signalling (Kotiaho [2001: 371]). The case for costs is most often indirect. This is because demonstrating that expenditures lead to a decrease in some component of fitness is never straightforward, even in organisms far simpler than humans. A brief digression about a costly signal in non-humans will make this point vivid.

Eyespan in male stalk-eyed flies (*Cyrtodiopsis dalmanni*) is plausibly a costly signal (Maynard-Smith & Harper [2003: 33-34]). Females prefer to mate with males with a wide eyespan. Males with certain genotypes develop a wide eyespan even on poor diets, while males with other genotypes develop a wide eyespan only when provided a high quality diet. Under conditions of scarcity, it appears, some males cannot afford the expenditure required by a wide eyespan. Furthermore, increased eyespan width has been shown to correlate with reduced overall aerial performance (Swallow *et al.* [2000]). Even given all this, the most that can be said about costs is that eyespan negatively affects something – aerial performance – that is most probably an important component of overall fitness.

Searcy & Nowicki (2005: 211-212) take a rather sanguine view of Kotiaho’s criticism of work on costly signalling. They admit that many empirical investigations of costly signalling assume that expenditures imply fitness costs. They consider the assumption warranted, however. Organisms have limited resources, time, and energy, and spending in any one area is bound to result in skimping in others. Searcy & Nowicki think Kotiaho’s objection does not significantly undermine empirical work on costly signalling, though they admit it highlights an important respect in which such work could be improved.

While Searcy & Nowicki’s reply may be warranted in some cases, it is a poor one in the case of costly signalling and human moral behaviour. For starters, it is harder to make a plausible inference from expenditures to costs to some component of fitness in the case of human behaviour than it is in relatively simpler cases like that of stalk-eyed flies. Expenditures will be
linked to such costs (if at all) more circuitously than, for instance, eyespan links up with viability via aerodynamic impairment. Also, the inference from expenditures to decreases in some component of fitness in the case of human moral behaviour is complicated by fact that expenditures on moral behaviour may have a positive effect on some components of fitness, via kin selection and reciprocity for instance. This means that assessing the impact on fitness of expenditures in the case of moral behaviour must be a matter of balancing the behaviour’s positive and negative influences on various components of fitness.

This problem does not arise in simpler cases like that of the stalk-eyed fly: there are no benefits to flying clumsily *per se*. Kotiaho’s objection to costly signalling claims must therefore be taken seriously by a proponent of SSM. Thus, as I work through the expenditures required by the various kinds of signalling behaviour mentioned in SSM, it should always be kept in mind that these expenditures provide at most plausible grounds for supposing the signals to be costly with respect to some component of fitness. Indeed, in some cases, as Miller fails to notice and as I will highlight, we have not even that much.

I will begin where the prospects for substantiating a costly signalling story seem most promising. Giving gifts and making charitable donations obviously requires expenditures. Both these behaviours require expending resources (production expenditure). Obtaining those resources in the first place may require time and energy and, depending on the particular resources, perhaps some degree of risk (developmental expenditure). I say ‘may’ because wealth can be inherited rather than earned: the entrepreneur and the trust fund baby both have a million dollars, the former by his own labour and the latter by that of his mother. The social conditions allowing for the accumulation and inheritance of great wealth are, evolutionarily speaking, recent enough that this complication can be left to one side. Miller’s claim is about the evolutionary history of our tendencies to give to others and to respond in particular ways to generous givers. Typically, among humans, generous gift giving would have required developmental as well as production expenditures. These expenditures can be plausibly linked to fitness costs, especially if we focus on the conditions under which humans evolved rather than those under which most of us currently live. Giving away a laboriously-gathered armload
of berries to support a sick neighbour during their illness, for instance, or intently scouring the seashore for a single splendid shell to delight a potential lover, would reduce the giver’s resources and expose the giver to risks in a way that would quite plausibly reduces the giver’s chances of survival.

Complications arise when we ask whether gift giving and donating resources to others are differentially costly behaviours, that is, behaviours that are more expensive to perform for individuals lacking the quality signalled. When the quality signalled is wealth, the case for differential costs looks promising. A given quantity of resources represents a greater proportion of a poor individual’s total than it does a rich individual’s total. Thus, the relative cost of giving away any particular amount of resources is higher for the poor individuals than for rich ones. When what is signalled is supposedly psychological qualities – altruism, kindness and sympathy – the case for differential costs looks shaky. Take two individuals of identical resource levels (we can imagine them to be Wall Street brokers with millions of dollars or prehistoric flint knappers with a stack of stone axes). One of our two individuals is a genuinely altruistic, kind, sympathetic soul who gives away a quarter of his resources to others in need. The other is image-conscious and self-aggrandising but likewise donates a quarter of his resources to the needy. The problem for SSM as Miller presents it (especially in his later [2007] work) is that giving away resources is not more costly for the image-conscious, self-aggrandising giver than for the genuinely altruistic, kind, and sympathetic one. If the cost of giving the signal is the same for both individuals, then it is hard to see, against the background of a Zahavi handicap understanding of costly signalling, how the signal could reliably reveal the kind of psychological traits Miller mentions.

Moving on, heroism as Miller understands it (i.e. endangering oneself to save others) looks like a good candidate for being a Zahavi handicap advertising physical quality. The expenditures involved are the risks posed by burning buildings, freezing ponds, hungry beasts, and the like. Those risks count as production costs in the scheme above. They can quite reasonably be taken to reduce the hero’s chances of survival. Further, it is not unreasonable
to think that those risks are even greater for a fat and sickly individual than a fit and healthy one. Again, though, when the behaviour is taken to be a signal of psychological qualities, things get murkier. Imagine a case that parallels the one laid out above: two equally physically capable individuals risk their own lives to rescue others. The first is genuinely concerned for those others’ well-being and courageous enough to act despite the danger. The second is a reckless thrillseeker and a gloryhound who performs the rescue solely to gratify desires for excitement and fame. As above, the supposed signal of psychological quality is equally costly for both individuals, making a Zahavi handicap-style costly signalling claim problematic. The general point here is that, in claiming heroism is a Zahavi handicap, care must be taken to specify the qualities such behaviour is supposed to signal, since different specifications (e.g. physical condition, altruistic concern, courage) will make the claim more or less plausible.

The problems for SSM get worse when we closely examine the claim that behaving in paradigmatically kind, sympathetic, fair, honest, conscientious, and patient ways serves as a costly signal of those psychological qualities. Consider fair behaviour first. Behaving fairly will occasionally mean forgoing the benefits of cheating, which can be counted as an opportunity cost paid to produce the signal. But whether fair behaviour is on balance costly will depend on a complex interaction between the benefits of cheating, the chance of detection, the probability that punishment will follow detection, and the severity of any anticipated punishment. In some, perhaps many, cases, behaving fairly may be the least costly option. Moreover, behaving fairly is most likely to be on balance costly when others are not around to detect cheating, but that is when the value of the behaviour as a signal is most dubious, due to the very lack of an audience. Much the same things might be said about behaving honestly.

When we consider sympathy, empathy, and kindness, the problems for SSM continue to mount. Miller distinguishes between sympathy and empathy as follows: “empathy is the mental capacity to understand the suffering of others” while “sympathy is the emotional capacity to care about [others’]
suffering” (2000: 330). As far as I can see, behaving in a way that shows one understands and cares about another’s suffering just is behaving kindly. Such behaviour is not very demanding in terms of expenditures. Granted, time spent “listen[ing] to sexual prospects enumerating their past sufferings” (Miller [2000: 330]) is time taken away from other activities, so behaving kindly does incur an opportunity cost. Also, listening to a sexual prospect’s tales of woe is no doubt boring and annoying for one who is not genuinely sympathetic. Doing so does not, however, require a great amount of resources or energy; neither does it typically involve much risk. Nor is such behaviour, to the extent that it is costly at all, differentially costly for genuinely kind individuals and fakers. This is not to say that nothing serves to maintain the reliability of such behaviour as a signal, only that costs may not be the place to look.

A costly signalling story a la Zahavian handicaps seems so inapt in the case of sympathy, empathy and kindness that it is worth looking to see what Miller has to say about the way in which such behaviours are meant to serve as signals. In his updated (2007) presentation of SSM, Miller writes that “moral virtues may function as good genes indicators by being difficult to display impressively if one has a high mutation load that impairs the precision of body and brain development” (2007: 101). By way of specific example, he suggests that “a conspicuously expert level of empathy may function as a sort of neurogenetic warranty” because the capacities needed to accurately gauge others’ mental states are complex and sensitive to disruption, as in autism for instance (2007: 101).137 Again, a detour through a non-human example will help here, since expert empathy as a kind of quality guarantee seems less like a handicap such as the peacock’s tail and more like a paradigmatic case of an index, namely, the courtship dance of the male fruit fly.

During courtship, the male fruit fly (Drosophila melanogaster) must manoeuvre to remain facing the female as she moves rapidly around him. If the male gets out of step in this dance, the female departs. Dancing ability

137 Notice, this kind of explanation would not be specific to moral virtue. It could be offered for any complex socio-psychological capacity, for example, the ability to tell jokes well. This is not something likely to trouble Miller (who in The Mating Mind proposed sexual selection as part of the explanation for humour), but it is worth noting.
depends on visual acuity and coordination. Genetically defective males have impaired vision and coordination and thus are clumsy dancers, making dancing ability a reliable — because unfakable — signal of genetic health. Notice, it is not possible to recast this case in terms of developmental costs rather than constraints. It is not the case that clumsy males could have been better dancers had they only invested more in developing their visual and motor systems. Rather, those systems simply will not develop properly and precisely unless an individual is free from mutations.

Supposing Miller is right about expert empathy serving as a "neurogenetic warranty," the way to understand empathetic behaviour as a signal would then be as an index rather than a handicap, since barriers to faking the behaviour rather than the costs of the behaviour would be what keep the signal reliable. Notice, genetic defects are only one possible barrier to faking. The difficulty of convincingly simulating certain psychological states is another (cf. Frank [1988]; see also §1.2.2c).

When it comes to psychological traits like conscientiousness and patience (Miller [2007: 101]), the associated behaviours often bring direct benefits. If these things are to be included in SSM, at a minimum it is necessary to distinguish between conscientiousness and patience in general and these traits when directed at another person. In general, being conscientious and patient while going about one’s daily tasks is beneficial. Being conscientious and patient with other individuals, though, does incur some costs, even if only time-based opportunity costs. As with kindness, however, there is little reason to think those costs are extremely high and little reason to think those costs fall differently on individuals who really possess the relevant psychological traits and those who merely pretend.

Let me pause and sum up the discussion so far. The moral behaviours most amenable to interpretation in terms of the Zahavi handicap version of costly signalling theory are those that advertise wealth or physical quality. It is difficult to maintain a Zahavi handicap-style story about psychological qualities like kindness, sympathy, fairness, honesty, and patience. The relevant
signalling behaviours often do not seem highly costly nor, when they are costly, do the costs seem appropriately different for honest versus dishonest signallers. This is a serious problem for Miller. In his updated version of SSM, a central claim is that moral behaviours signal psychological qualities, and Miller remains focussed on costly signalling:

[A]ll moral virtues displayed during courtship are potentially fallible signals of other traits or future traits, so their reliability and stability must be analysed in a costly signalling framework (2007: 99).

The final thing to discuss is the plausibility of public moralising playing a costly signalling role. As mentioned above, Miller mentions this idea but does not discuss it in detail. I suggested that if we are interested in understanding human morality as a kind of self-advertisement, then public moralising is worthy of attention. After the discussion above, however, the problems with the idea that public moralising can serve as a Zahavi handicap-style costly signal are probably already apparent. In everyday contexts, the expenditures required by public moralising are minimal. In some circumstances, of course, outspoken moralising might involve significant risks, but protesting against ruthless regimes, for example, or volubly damning dangerous anti-social elements of society should be classed as heroism of the Rosa Parks kind. In everyday contexts, public moralising is simply not costly in the kind of way required by the Zahavi handicap version of costly signalling theory.

If I have done what I set out to do in this section, things will seem grim for SSM. Only some of the purported signalling behaviours have turned out to involve expenditures that can be plausibly linked to costs in the way required by the Zahavi handicap version of costly signalling theory. Other of those behaviours seem more like indices than handicaps, suggesting that constraints (as opposed to costs) should play a bigger role than Miller recognises in a signalling-based account of morality (as they do in Frank’s [1988] account of emotions). Yet other purported signalling behaviours seem neither prohibitively costly to fake nor constrained to be honest, leaving it mysterious
how their reliability is maintained, if indeed a signalling account is appropriate there at all.

4.3. Beyond the Handicap Principle

In this section, I will propose a fix for SSM. Many of the problems laid out in the previous section stem from basing SSM on an insufficiently sophisticated understanding of the ways in which costly signalling can work. The claim that certain behaviours — fair, kind, honest, patient, and conscientious behaviours, as well as public moralising — signal psychological qualities was particularly hard to fit into the Zahavi handicap version of costly signalling theory. In what follows, I will attempt to show how these problem cases can be accommodated within a more sophisticated costly signalling framework.

There are three pieces to put in place. The first has already been placed: it is the distinction drawn above between intrinsic and imposed costs. Placing the next piece will require detailing the developments in costly signalling theory that have been made since Zahavi first proposed the Handicap Principle. The final piece is yet another distinction, this one between signals of quality and signals of intent. With these three pieces, I will make the case that the problematic behaviours listed above can actually be understood in terms of costly signalling.

4.3.1. Costly Signalling Theories

Zahavi (1975) saw the costs of signalling as imposing a kind of test on signallers and thought that if those costs were high enough, then the test would be difficult enough that only genuinely high quality individuals could pass it. The challenge, Zahavi thought, was to survive despite having paid the high costs of signalling. By paying the high costs of signalling and still surviving, a high quality individual proves to others that s/he is indeed high quality.

Zahavi later refined his ideas about how costly signalling worked, suggesting that “the phenotypic manifestation of the handicap is adjusted to
correlate to the phenotypic quality of the individual” (1977: 603). The crucial difference between this version of the costly signalling idea and Zahavi’s initial version is as follows. In the initial version, high quality and low quality individuals are assumed to pay the costs of signalling and only high quality individuals are assumed to survive (at least for very long). In the later version, high signalling costs are paid only by those who can afford those costs, that is, the genuinely high quality individuals. Individuals who cannot afford high signalling costs either do not signal at all, or, signal at a lower intensity that is affordable given their quality. Signals in former case have been termed “conditional” handicaps, and in the latter, “revealing” handicaps (Searcy & Nowicki [2005: 10]).

The conditional handicap and revealing handicap versions of costly signalling theory preserve intact Zahavi’s view that honest signals must be costly. A more radical reworking of costly signalling theory has recently been offered by Lachmann et al. (2001). Lachmann et al. believe Zahavi’s influential first formulation of costly signalling theory has led later researchers to focus too intently on the idea that honest signals must be costly. This has led to neglect of what Lachmann et al. believe is the key insight in the costly signalling idea. That insight, they say, is that if signalling is “differentially costly according to the meaning and/or validity of the message [then] signallers can be relied upon to report aspects of their condition or of the environment” (2001: 13190). Crucially, they say, this condition can hold even when honest signals are cheap or free, provided that dishonest signalling is prohibitively costly.  

Lachmann et al. (2001: 13190) offer a toy example to illustrate their point. Imagine we have signallers and receivers and that signallers vary with respect to some quality of interest to receivers. Imagine that the higher a signaller’s quality is perceived to be by receivers, the better off that signaller is. Imagine that receivers benefit from accurately assessing signallers’ quality.

---

138 See Bergstrom et al. (2002) for a more mathematically challenging presentation of the same ideas.

139 Diego Gambetta (2009) provides an entertaining discussion of this idea in a human context, that of communication and trust among criminals.
Finally, imagine that signalling one’s quality up to its true level is free but any exaggeration is lethal. Under these conditions, signallers would reliably signal their true quality and receivers would trust those signals, not because honest signals are costly but because dishonest ones are.

In the case of the peacock’s tail, the differential cost requirement for reliability is met because signalling a given level of quality is costly for individuals who honestly are of that quality but even more costly for individuals who dishonestly pretend to be. In the case just given, that requirement is met because signalling a given level of quality is costly for individuals who dishonestly pretend to be of that quality but not costly at all for individuals who honestly are. In both cases, though, the crucial condition is met. Contra Zahavi, then, honest signals need not be costly in order for reliability to be maintained in signalling systems, even when signaller and receiver interests conflict.

While the cost of dishonesty is unlikely to be death in many real-world cases (as it is in the toy example above), such a drastic cost to dishonesty does serve to emphasise the point. Lachmann et al. also describe a less extreme scenario based on a real-world example (2001: 13191). Male sparrows use minor variations in plumage (such as different throat bibs or forehead patches) to signal their fighting ability. It is almost free for a genuinely strong and aggressive sparrow, which is able to win fights, to signal that he is such, since all he need do is grow the dominant ‘badge.’ A weakling sparrow that grows those same (cheap to produce) feathers will, however, be soundly thrashed when challenged by other males. Signalling one’s true quality in this situation is virtually cost-free, but dishonest signallers pay for their pretence.

4.3.2. Signals of Quality vs. Signals of Intent

There is a distinction drawn in the biological literature on signalling between signals of quality and signals of intent. Signals of quality provide information about an extant feature of the signaller, for example, its size, hunger level, or resistance to parasites. Signals of intent, by contrast, “convey messages about
contexts that have not yet occurred at the time the signal is produced” (Laidre [2009: 337]). For example, a signal of intent may provide information about how the signaller is likely to behave in future. Signals of intent have received less attention than signals of quality (Gangestad & Thornhill [2007]; Gurven [2004]). Such signals have, however, been intensively studied in the context of aggressive interactions, such as disputes over resources like food or territory.¹⁴⁰

The prospects for fitting signals of intent into a costly signalling framework may seem poor. Signals of intent typically involve low intrinsic costs. For example, Enquist et al. (1985) suggested that, among Northern fulmars (Fulmarus glacialis), adopting a breast-to-breast posture with a rival may signal intent to fight, since that posture is highly effective in making rivals retreat during contests for food. Little time or energy is required to adopt the breast-to-breast posture. However, Enquist et al. also noted that adopting that posture was likely to land an individual in a fight. Enquist’s (1985) modelling of signals of aggressive intent demonstrated that such signals could be reliable, provided that signallers could not signal intent to fight then escape if their rival attacked. That is, the signals could be reliable so long as they carried imposed costs in the form of increased risks of attack and/or injury. Adopting the breast-to-breast posture requires the signaller to be very close to his rival, meaning that escape may be difficult if that rival decides to attack. Enquist’s condition may well be met in this case. Maynard-Smith & Harper (2003: 18-19) and Searcy & Nowicki (2005: 144-146) both consider the claim that signals of aggressive intent involve such costs to be plausible, if currently empirically understudied.

¹⁴⁰ Early work reached sceptical conclusions about the role of threat postures in signalling the intent to attack, since the adoption of such postures was not reliably followed by actual attack (Stokes [1962]; Caryl [1979]). Later work by Magnus Enquist ([1985]; Enquist et al. [1985]) recognised that signallers’ behaviour after signalling was likely to be contingent upon receivers’ responses. Enquist suggested that threat posturing signalled, not the intent to attack regardless of whether a rival retreated but, rather, the intent to attack if the rival did not retreat. Data like that of Stokes does not undermine the idea that threat postures serve as signals of intent in this conditional sense. After all, some rivals will retreat in the face of a threat and actual attack will be unnecessary.
The problem cases for SSM identified above – kind, fair, honest, conscientious, and patient behaviour, and public moralising – may be better understood as signals of intent than as signals of quality. The suggestion that kind, fair, etc. behaviour serve as signals of intent is consonant with some of what Miller has to say in his presentations of SSM. Miller notes that “whenever there are incentives to act like a better partner during courtship than after reproduction, the problem of trait stability arises” (2007: 101). The “problem of trait stability,” as Miller sees it, is that individuals may behave kindly, fairly, etc. when signalling to potential mates and may really be kind, fair, etc. but may lose those qualities at some later point: “the present value of a trait is unreliably correlated with its future value” (2007: 99). Good mating prospects might turn bad. This is certainly one way to see things. I am suggesting that things can be seen slightly differently, the question being not whether the quality reliably signalled now will still be possessed by the signaler later, but whether signalers can be relied on to follow through on the intentions they signal now. Bad mating prospects might seem to be good ones.

Regardless of how well the suggestion that kind, fair, etc. behaviour serve as signals of intent fits with Miller’s presentation of SSM, the suggestion is, I think, a plausible one. When a female $A$ observes a potential mate $B$ behaving kindly (etc.), the question of fundamental importance to $A$ is whether $B$ would behave in those ways toward her (and their offspring) should she pair up with him. When $A$ moralises in a certain way, say lauding helpful behaviour and deploring cutthroat selfishness, what $B$ wants to know is whether this pronouncements predict the way $A$ herself would act in a partnership with him.

4.3.3. Revising SSM (again)

According to Lachmann et al., the crucial requirement if costs are to underwrite reliability in signalling systems is that it be more costly for a signaler to signal dishonestly than honestly. According to the suggestion made in the previous section, kind, fair, honest (etc.) behaviours signal agents’ intentions to act in kind, fair, honest (etc.) ways in future. According to the
classification scheme laid out earlier in this chapter (see §4.2.1), one class of costs is imposed costs, those that result from actions taken toward the signaller by receivers and observers of the signal. Putting these three elements together allows, I think, a case to be made that kind, fair, honest (etc.) behaviours, along with public moralising, do fit into a costly signalling framework, as Miller claims but fails to show.

Once again, I will structure the discussion to come by posing some questions. First, are dishonest signals of intent to be kind, fair, honest (etc.) costly? Second, is it more costly for a given individual to signal dishonestly than honestly about their intentions to be kind, fair, honest (etc.)? I think the answer to the first question is pretty clearly 'yes.' I think the answer to the second question is less clear but there are things to be said in favour of a positive answer.

The importance of reputation in human social life is the place to start in laying out the costs of dishonest signals of intent to behave kindly, fairly, and so on, as well as the costs of hypocritical moralising. A person who behaves kindly, fairly, and so on when under observation and is later caught behaving cruelly, unfairly (etc.) might be assaulted as are birds who dishonestly signal their intent to attack then try to flee. This kind of brute physical punishment is probably not the primary cost of dishonesty in the human case, though. The social ramifications of being caught misrepresenting one’s intentions are probably far more important. As Maynard-Smith & Harper note:

In social interactions, the honesty of signals can be ensured by ‘reputation’: there may be an immediate advantage from giving a dishonest signal, but this is more than counterbalanced in future interactions if the signaller acquires a reputation for dishonesty (2003: 121).

Maynard-Smith & Harper identify several ways in which reputation effects might help enforce honesty in signalling systems (2003: 121-122). In the case of “direct reputation,” $A$ deceives $B$ with the result that $B$ distrusts $A$ if they interact again at some later time. In “indirect reputation,” $C$ observes $A$

Maynard-Smith & Harper rightly point out that the operation of direct, indirect, and reported reputation effects requires increasing cognitive sophistication on the part of receivers. The cognitive demands of reputation tracking limit the role of reputation effects in stabilising biological signalling systems (cf. Stevens & Hauser [2004]). However, even if many non-human organisms are not are up to the task, humans certainly manage that feat. Indeed, human social life is structured in a way that makes reputation effects especially powerful. As Searcy & Nowicki note, the effectiveness of reputation effects in securing reliability is greatest when signalling occurs in "communications networks" that allow "eavesdropping," that is, when:

...individuals not directly involved in a signalling interaction [can] observe both the signals given and the interaction outcome, learn something about the reliability of one or both the interactants, and use this information in shaping their own subsequent responses to those signallers (2005: 218).

The kind of conspicuously helpful behaviours Miller claims play a signalling role in the context of mate choice are performed within just such networks. For this reason, I think it is reasonable to claim that the risk of ruining one's reputation should be counted as a cost of dishonestly signalling the intent to act fairly, kindly, and so on. Fakers who are discovered may suffer a range of bad consequences, from reduced opportunities to interact profitably with others all the way to total ostracism.

---

141 One might wonder at this point whether a signalling story is necessary to explain helpfulness, cooperation and so on in the kind of social groups humans have inhabited for the vast majority of our evolutionary history. Recall the distinction between signals and cues (see §1.2.2c). Cues include any aspect of the world that provides information that can guide future behavior. In small-scale, low-mobility social worlds, a great deal of relevant information about potential partners will be available solely via cues, 'leaking out' (so to speak) from such individuals' ecological and economic interactions with others (see Sterelny [2007: 726]). While it is true that cues provide much useful information about potential partners, this should not lead us to think a signalling story is otiose. One may have a quality product and rely on word of mouth to make that fact known, with some sales success. A sales gain may still be made, though, by actively promoting one's product, that is, by advertising its quality. Signaling in SSM is essentially advertising. The fact that cues provide information about one's quality as a partner need not mean a signalling story is redundant, so long as signalling can gain one entry into more and/or better partnerships than relying on cues alone.
Obviously, fair and kind and honest behaviours are costly in the way just described for dishonest signallers but not for folk who are genuinely fair, kind, and honest. More than this must be said, though. The case must also be made that signalling dishonestly doesn’t pay, that is, that is it more costly for a given individual to signal dishonestly than honestly about their intentions to be kind, fair, and so on.

That dishonesty doesn’t pay when it comes to signals of kind, fair (etc.) intent may at first seem like a tough case to make. Consider: if one dishonestly signals kind, fair, honest (etc.) intentions to potential mates and is caught out, one’s reputation will suffer, but if one honestly signals one’s intent to cheat and malinger in a partnership, one’s reputation will certainly be damaged and one’s mating chances plummet. For unkind, unfair folks, deception at least has a chance of scamming a reproductive partner, whereas honesty seems tantamount to declaring mating-market bankruptcy.

Things are not so grim, however. I have just compared the consequences of dishonestly signalling kind (etc.) intentions with those of honestly signalling that one is, in fact, a bit of a bastard. However, those are not the only options. We should include in the comparison the consequences of such folk devoting their signalling effort to honestly advertising attractive qualities they really do possess. Individuals in the mating market will have varying strengths and weaknesses. It makes perfect sense that those weak in one area would try to cover that fact. This need not be done by lying about it. Rather, it may be done by distracting from it with loud announcements of strengths elsewhere (material resources, maybe, or physical condition).

If we consider the right range of options, it is plausible that signalling honestly is a better bet than signalling dishonestly, at least for many signallers much of the time. Of course, this is an empirical matter. I am concerned here not to settle it but only to clarify how a costly signalling story about kindness (etc.) would have to go to be plausible. As Miller presented SSM, the claim that behaving kindly, fairly, honestly, and so on, was a form of costly signalling was deeply implausible. I will be content if, in what has been done
above, it is now clear how those behaviours could be plausibly fit into a costly signalling framework.

To close this section, I will note and discuss Maynard-Smith & Harper's reluctance to count signals that are kept reliable by punishment as costly signals. They write:

There is one context in which we prefer not to call a signal a handicap, although it may have costly consequences. Suppose there is 'punishment'. An individual indicates that it will do A, but then does B, and is punished by the receiver: lying signals are costly. Our reason for not referring to such cases as handicaps is that at an evolutionary equilibrium, lying will be rare or absent, and signals can be both cost-free and reliable. The problem lies, not in explaining reliability, given punishment, but in explaining why the receiver is willing to incur the cost of punishing (Maynard-Smith & Harper [2003: 8]).

The first thing to note in considering Maynard-Smith & Harper's comments about punishment is that it is not costly consequences per se that they are reluctant to include in costly signalling. They consider behaviours such as fulmars' breast-to-breast display (described in §4.3.2) to be costly signals and, more generally, they count predation risk as one of the kinds of costs that can underwrite reliability (2005: 7, 8). Their problem is rather with taking costs resulting from punishment of dishonesty as an explanation for why signallers are honest. As Maynard-Smith & Harper rightly note, this kind of explanation for signal reliability is importantly incomplete. That receivers and observers of dishonest signals sometimes take the time, spend the effort, and run the risks of dishing out punishment is a puzzle in its own right.

That individuals engage in costly punishing behaviour certainly requires an explanation. It has been shown that punishment of free riders – those who take the benefits of cooperation within a group but do not contribute to the production of those goods – can lead to the maintenance of cooperative behaviour in the group (Boyd & Richerson [1992]). There remains the issue of what stabilises punishment, though: a so-called 'second order' free rider problem exists. Insofar as punishment is costly to the punisher but benefits other members of the group (by weeding out free riders), punishment itself can be seen as a kind of cooperation. Punishment itself thus stands in
need of explanation, and an account of cooperation in terms of punishment is incomplete without that further explanation.

The same kind of problem exists for punishment as a means of stabilising reliability in a signalling system. I agree entirely with Maynard-Smith & Harper's point that, insofar as punishment of dishonest signals is costly not only to the one punished but also to the punisher, some further story is needed. In fact, for imposed costs in general, if imposing the cost is itself costly, then we need an account of why receivers and/or observers of dishonest signals would pay those costs. However, I see this as a demand to be met rather than as something that disqualifies such signals from counting as costly signals. The question is whether the right kind of further story can be provided. I think it can.

To begin with, it is worth restating that the kind of punishment I think is important in discouraging dishonest signals of kind, fair (etc.) intent is not the infliction of brute physical damage in face-to-face confrontation with fakers. While this kind of punishment may occur, the primary mode of punishment is likely to be social: reputation damage, for example, and ostracism. Such socially-mediated punishment can be meted out by the simple expedient of talking about fakers. Indeed, one need never even encounter the individual that one is punishing. It is tempting to see the costs of punishing those who dishonestly signal kind, fair (etc.) intent as quite low. Complicating matters is the fact that gossip often brings with it the risk of making enemies and losing allies. Such risk must be factored into the cost of socially-mediated punishment. Even so, that risk is quite different from the obvious and high risks of physical confrontation that make Maynard-Smith & Harper leery about using costly punishment to explain signal reliability.

The next thing to say is that even if punishing is in some ways costly, there may be a benefit to the punisher that outweighs those costs, making punishment something that is, overall, to the punisher's advantage. Miller makes an interesting suggestion along these lines. SSM itself, he suggests, might help solve the second order free rider problem. He writes:
[SSM] identifies selfish mate choice incentives (e.g., good gene and good parent payoffs) for “rewarding” the virtuously punitive with sexual relationships. That is, the effective imposition of punishment on antisocial others (at substantial risks and costs to oneself) should be seen as virtuous, socially status-enhancing, and hence sexually attractive (2007: 111).

Costly punishment might itself be understood as a costly signal. By meting out costly punishment to those who pretend to be kind, fair, honest, and so on, and to those who hypocritically moralise, punishers may effectively advertise attractive qualities. Most straightforwardly, punishing might display strength and bravery. More abstractly, it might communicate the punisher’s commitment to the transgressed ideals of honesty, fairness, and so on.

Of course, these claims about a signalling role for punishment are speculative. A costly signalling theory of punishment would have to be developed with all the many cautions and clarifications made throughout this chapter in mind. I will not be doing that here. I will, however, admit to finding pleasing the possibility of a neat dovetail between costly signalling, punishment, the problem of reliability, and the second order free-rider problem. I also find it promising that empirically-oriented researchers, specifically, anthropologists Eric Smith & Rebecca Bliege-Bird (2005), have recently considered the possibility that punishment might play a costly signalling role. Smith & Bliege-Bird offer only a verbal model of how they think punishment as a costly signal might work (2005: 135). They do not provide any direct empirical evidence for the hypothesis. Research aimed at finding such evidence would be very welcome. For the moment, I will have to be content with some plausible speculation.

Recall, part of the reason that Maynard-Smith & Harper are reluctant to count signals that are kept reliable by punishment of dishonesty as costly signals was that “at an evolutionary equilibrium, lying will be rare or absent, and signals can be both cost-free and reliable” (2003: 8). It is worth noting that Searcy & Nowicki share Maynard-Smith & Harper’s reservations on this point. Searcy & Nowicki consider the idea that signal reliability can be maintained by the costs of being punished for dishonesty to be “so far from the original spirit of the handicap principle, in which signals produce costs
because of their extravagance, that [this] should be regarded as a separate hypothesis” (2005: 217).

Zahavi’s original version of costly signalling theory did focus on the high costs of extravagant signals like the peacock’s tail. Only the most dedicated devotee of costly signalling theory – Zahavi himself – still thinks that costs are always required to maintain reliability in signalling systems:

[T]he handicap principle is effective not only in mate choice, but in the selection of all signals, that is, every signal, sexual or other, should contain handicaps to enable the receiver to ascertain its reliability (Zahavi [2007: 144]).

Costly signalling theory has itself evolved over time. It is not clear to me that staying true to Zahavi’s initial idea is what theorists working in this field should try to do. Zahavi’s insight was that signal reliability can be maintained if dishonesty, while possible, is not profitable. Work like that of Lachmann et al. refines rather than replaces that insight.

4.4. Conclusion

This chapter has drawn and deployed numerous distinctions in order to develop a conceptually cleaned-up version of SSM. According to my cleaned-up version of SSM, the signals in the case of morality are various actions: giving goods away, running risks to rescue others, lending a sympathetic ear or a shoulder to cry on, doing at least one’s share of the work and taking no more than one’s share of the profits, telling the truth, and publicly pronouncing on others’ moral fortes and foibles. The qualities signalled by these actions are the material resources and/or physical condition and/or psychological states of the agent. These qualities in turn serve as cues in mate choice, that is, features it pays to pay attention to when choosing a mate.

This chapter also attempted a scrupulous accounting of the costs of moral behaviour. The upshot of this accounting was that, for some of the moral behaviours that Miller claims are signals, the costs of dishonesty seem
less important in underwriting reliability than do constraints against faking. This finding nicely links SSM to Robert Frank's work on strategic role of the emotions in signalling commitment. Further work on the signalling role of moral behaviour could usefully be pursued in the context of Frank's ideas.

I take the finding that certain moral behaviours are more likely to be indices than handicaps to represent a kind of progress for SSM. Even so, I have tried in this chapter to show how a sufficiently nuanced understanding of costly signalling can accommodate some of these apparent problem cases. Miller presents SSM in terms of a rather simplistic version of costly signalling theory, closely tied to Zahavi's initial presentation of the idea. I have presented a more sophisticated take on that idea. According to this take, the key insight behind costly signalling theory is that reliability can be maintained so long as dishonest signalling is prohibitively costly, which is a condition that can be met even when honest signals are cheap or free. I have suggested that this understanding of costly signalling theory can help us understand how public moralising -- an important aspect of our moral lives -- can serve as a reliable signal, despite not being constrained to be honest and not being, intrinsically at least, a costly activity.

Much in the way of empirical data is yet to be provided in support of the claim that moral behaviour plays a signalling role in the context of mate choice. It is my hope that my discussion in this chapter might provide a framework to guide attempts at gathering such data. Getting clear on the supposed signals, the qualities supposedly signalled, the costs involved in signalling, and the role of constraints in underwriting honesty seems to me a prerequisite for doing fruitful empirical investigation into the signalling role of morality.
Chapter 5. SSM: Empirical Evidence

5.1. Introduction

This chapter will evaluate the evidence for SSM, and will conclude my efforts to clarify, elaborate, and defend Miller's suggestion that the evolution of human morality might be explained in terms of sexual selection and costly signalling.

My first task in this chapter will be to clarify the empirical commitments of SSM in order to forestall some misguided objections to the view. I will clarify SSM's predictions about the development of moral capacities over life history stages, and about sexual dimorphism in human morality. I will also consider empirical work that bears on those predictions.

Next, I will turn to empirical work relevant to SSM. I will not be producing any new data, being as I am firmly ensconced in my philosopher's armchair. Instead, I will survey several fields - psychology, anthropology, and experimental economics - and synthesise findings that offer some support for SSM. The findings of psychologists, anthropologists, and experimental economists support SSM, I will show, insofar as they indicate that behaving fairly, generously, honestly, and so on, serves, in some circumstances, as a signal.

Finally, I will present what I call the 'many receivers' problem for SSM. The problem is that current empirical work suggests that the signalling role of moral behaviour is not confined to attracting mates but extends to obtaining cooperative partners more generally. In light of these findings, SSM appears to be part, but only part, of a more general signalling-based account of morality. Having presented the many receivers problem for SSM, I will close by highlighting the contribution that sexual selection in particular can make to a more general signalling-based account of moral judgement and behaviour.
5.2. Clarifying the Commitments of SSM

In this section, I will do some groundclearing work. In my experience, there are several recurrent but misguided objections that tend to be raised to the claim that our moral capacities are the products of sexual selection. Although I have substantially reformulated Miller's presentation of SSM, my version of the view as much as his must deal with these objections. In developing my version of SSM, I have not made any predictions regarding the issues to be raised in this section. Miller has, though, and I will take those as starting points in working out the best position to adopt on those issues.

One of the objections I will consider in this section can be deflected rather easily but the others will require a little more effort to block. The heftier objections target SSM's predictions regarding, firstly, the development and expression of moral capacities across life history stages and, secondly, the existence and extent of sexual dimorphism in human morality. I will confront these two objections in the next couple of sections, after disposing of the more niggling objection below.\[142\]

The (in my experience) annoyingly common objection to SSM that I wish to get out of the way at the very outset of my discussion typically starts with a couple of pop culture proverbs: 'nice guys finish last' and 'good girls like bad boys.' One can, however, grant that displays of kindness, generosity, fairness, and the like are not attractive and may even be off-putting to some potential mates, without undermining SSM. Miller's brief comment that "to argue that moral virtues evolved through mate choice is not to argue that vice is never attractive" (2007: 99) is on the right track but needs some elaboration.

First, SSM is a claim about the evolutionary history of certain species-typical aspects of human psychology. As such, it does not make claims about the particular psychological make-up of every individual everywhere. It is

\[142\] One thing I wish to set entirely aside is the situationist challenge to character (Harman [1999]; Doris [2002]). Given that the version of SSM that I refined in Chapter 2 is not, like Miller's version, stated in terms of virtue ethics, it is not necessary for me to enter the debate in moral psychology about whether character traits are global or local.
entirely compatible with SSM that environmental influences powerfully affect individuals' sexual preferences. To draw a pertinent parallel, just as the theory of kin selection is not undermined by intra-family feuds, so SSM stands even if some individuals do not find moral displays sexually attractive.

Second, to talk (as Miller does) about vice being attractive is perhaps to take the pop 'wisdom' too seriously. Certainly, some folks are attracted to individuals who lie, cheat, steal, and hurt others. We need not suppose it is the lying, cheating, stealing, and infliction of pain that are attractive, though. Instead, the confidence and competence displayed by such characters may be the attractive things about them.

To be fair, Miller makes something like this point in a discussion of SSM (2008: 223), noting that whether females prefer “niceness” or “dominance” in a mate is a complex question. The key point is the shift from talk of ‘bad boy’ types to talk of ‘alpha male’ types: the latter may not be nice, but they need not be vicious either.

This kind of response to the ‘bad boys’ objection is supported by recent research on mate preferences (Lukaszewski & Roney [2009]) that has revealed a level of complexity to the well-established preference for kindness that Miller so often references. Lukaszewski & Roney found that male and female subjects’ preferences for kindness in a potential mate are indeed very strong when that kindness is specified to be “self-directed,” which somewhat confusingly meant directed toward the subject and the subject’s friends and family (2009: 30). When kindness is specified to be “other-directed” (i.e. directed toward anyone aside from those just named), subjects expressed a somewhat weaker preference for kindness in a potential mate (2009: 31). Note, in this latter case, subjects did not disprefer the trait, nor was the preference for it weak, just weaker than in the contrasting, self-directed case. Finally, Lukaszewski & Roney found that the preference for other-directed kindness in a mate was roughly on par, strength-wise, with the preference for other-directed “dominance,” meaning “behaviours that employ forceful or competitive tactics in order to promote desired outcomes” (2009: 30).
Revealing this level of complexity in the often-reported preference for kindness in a potential mate serves, I think, to show how SSM can deal with the ‘bad boys’ objection.143

Finally, I cannot resist briefly mentioning some speculation by anthropologists Smith & Bliege-Bird (2005) about a possible costly signalling role for immorality per se. Smith & Bliege-Bird suggest that flouting norms of fairness, for example, might be attractive to potential mates if doing so serves as:

...a signal of social power or superior status, reliable to the extent that norm violation is more costly than simply “paying one’s share”, and hence not viable for low status or subordinate individuals (2005: 136).

While there is currently no empirical support for this idea, it is at least intriguing insofar as it offers a way to reconcile the attractiveness of (certain) vices per se with SSM.

5.2.1. Morality and Life History

The iconic sexually-selected trait that has been mentioned repeatedly throughout my discussion is expressed only at sexual maturity: a peacock first grows an elaborate tail when he reaches breeding age in his third year of life (Jackson [2006: 24]). It is worth asking whether SSM predicts that sexually-selected moral traits will be expressed only at sexual maturity. This question is worth asking because, quite obviously, pre- and post-reproductive individuals judge and behave morally. If SSM predicted otherwise, the hypothesis would be in serious trouble.

Miller claims that SSM does not predict that displays of kindness, honesty, courage, generosity, and the like will be entirely absent in pre- and post-reproductive individuals. According to Miller, though, SSM does predict a specific pattern in the display of moral qualities across life history stages. Moral displays “should be low before puberty, increase rapidly thereafter, then

143 See Birkhead (2002) for an accessible and entertaining account of the complexities involved in human and nonhuman female mate choice.
decline gradually” (2007: 113) because young adulthood is the time of peak mating effort.

Miller does SSM no favours here. Who, one wonders, could believe that teenagers and twenty-somethings are, compared to other age demographics, kind, generous, honest paragons of virtue? A crucial first move in maintaining the plausibility of SSM here is to distinguish between different forms of moral display. On the one hand, take risky displays of courage and, on the other, displays of kindness and generosity. If risky behaviour in general peaks in young adulthood, it may well be true that risky moral displays are most common during that life history stage, but this does not mean moral display as such is most frequent and intense during that period.

Whether Miller is correct about the pattern of moral display across life history stages depends on what sorts of moral behaviour we examine. We see plenty of kind, fair, honest, generous, and sympathetic behaviour displayed by young children and older folk. Notice, Miller says moral display in such age groups should be low, not just lower than in young adults. In the cases just mentioned, Miller’s prediction is not borne out.

If behaviours that, according to SSM, serve as signals to potential mates are performed by individuals either too young or too old to mate, this presents something of a puzzle for the view. The same can be said about the tendency to make moral judgements. Recall, on the version of SSM refined in Chapter 2 (see §2.2.3b), making moral judgements serves to motivate intersexually selected signalling behaviours. The results of the moral/conventional task (see §2.3.1) indicates that moral judgements are made by children as young as three years old. Some account must be given of how this fits with SSM. I think such an account can be given for the prevalence of moral judgement and behaviour in pre-reproductive individuals. I will lay out that account below. When it comes to post-reproductive individuals (or even those

---

144 Nobody who went to high school or college, I suspect, nor the parents of anyone who did.
already in parenting partnerships), I think things are less rosy for SSM, but I will defer that discussion until later (see §5.4).

Unlike the peacock’s tail, some sexually-selected traits, such as birdsong, see significant investment long before sexual maturity. Early investment in the development of traits that will affect reproductive success later in life has been termed the “accumulation of reproductive potential” (Geary [2002: 45]; see also Geary & Bjorklund [2000]). In light of this, the worry that SSM may predict the absence of moral displays during pre-reproductive life history stages betrays a certain biological naivety. Even Miller’s less extreme claim that moral displays will be low in intensity and frequency prior to reproductive maturity seems suspect.

It should strike us as silly to expect that young people will suddenly exhibit a complex array of new psychological capacities and behaviours the minute they start to fancy members of the opposite sex. As Aristotle claimed in his *Nichomachean Ethics*, developing moral competence and character is a matter of long practice and training. The development of moral competence long before sexual maturity and the prevalence of moral behaviour in the lead up to sexual maturity might be understood as pre-reproductive individuals being in training for courtship contests later in life. The exercise of moral judgement and behaviour by young children can be fruitfully compared with play in young predators such as lion cubs. In the latter case, it is not at all puzzling why cubs that are still suckling should engage in stalking behaviour: doing so prepares them for the time when they will have to hunt for their food. Likewise, even though young children do not need to attract sexual partners, moral development may begin early to prepare them for the time when they will have to do so. The early development of the tendency to make moral judgements and the prevalence of moral behaviour prior to sexual maturity in humans is thus, I would claim, entirely compatible with SSM, and Miller is mistaken in saying the view predicts low juvenile moral display.
5.2.2. Morality and Sexual Dimorphism

Sexual selection can lead to striking differences between males and females of the same species. The examples already given in Chapter 3 (see §3.3.1) will serve us again here. Male elephant seals are much larger than female elephant seals, peacocks bear brilliant plumage but peahens are drab, and male bowerbirds perform elaborate displays in front of their decorated bowers while females merely observe. Many of the most familiar examples of sexual selection are ones in which the process has produced sexual dimorphism. Even the cases upon which Roughgarden (see again §3.3.1) bases her objections to sexual selection theory are cases in which males and females differ as a result of the action of sexual selection, albeit ones where females rather than males bear ornaments or engage in showy displays.

It is worth asking whether SSM predicts sexual dimorphism in human morality. Quite obviously, both men and women judge and behave morally. If SSM predicted otherwise, the hypothesis would be in serious trouble.\textsuperscript{145} Miller is thus understandably at pains to say that “a sexual selection account of moral virtues does not imply that males evolve all the conspicuous virtues and females play the passive role of virtue-assessment” (2007: 102). Miller here denies that SSM predicts what I will call ‘stark’ sexual dimorphism in human morality, that is, the kind of sexual dimorphism in which one sex bears a sexually selected trait while the other sex bears only the preference for that trait and not the trait itself.

The reason, according to Miller, that a sexual-selection based account of morality need not imply stark sexual dimorphism is that “given mutual

\textsuperscript{145} Notice that if SSM predicted stark sexual dimorphism in human morality, the hypothesis could not be saved even by a so-called ‘hitch-hiking’ explanation (Symons [1979]; Lloyd [2005]). Some traits are sexually monomorphic but functional in only one sex. For example, nipples serve an important purpose for women but are non-functional in men. The reason both sexes have them is that nipples develop during gestation prior to sex determination and, crucially, impose no significant costs on the sex in which they are nonfunctional. Hitch-hiking explanations rely on the hitcher being relatively little bother. Morality, however, is a boisterous and smelly hitcher who demands to stop at every roadside attraction and explore every scenic detour. That is to say, a hitch-hiking explanation could not come to the rescue of SSM if SSM predicted stark sexual dimorphism in human morality, because the capacity for moral judgment and behavior is not a selectively neutral trait.
[mate] choice, both human sexes should show conspicuous, sexually attractive virtues during mate attraction and retention" (2007: 102). Mutual mate choice can indeed lead to monomorphic sexually selected traits. In socially monogamous, cooperatively breeding birds such as swans and auklets, for instance, mutual mate choice has produced sexually monomorphic ornamentation (Kraaijeveld et al. [2004]; Jones & Hunter [1993]). Moreover, humans are a species in which both sexes are choosy when it comes to mating (Geary et al. [2004]). It is worth pointing out that mutual mate choice alone will not lead to a lack of sexual dimorphism in the traits under sexual selection: both sexes must also be choosing based on the same criteria. I suspect that in the background of Miller’s discussion lurks Buss’ (1989) cross-cultural mate preference study. Recall, that study found that men and women alike judged kindness to be the most important feature in a potential long-term mate, ranking it above health, wealth, intelligence, and attractiveness (see §1.2.1). Collectively, the considerations above provide reasonable grounds for denying that SSM implies stark sexual dimorphism in human moral capacities, which is welcome news for the view.

There remains the question of whether SSM predicts what I will call ‘subtle’ sexual dimorphism in human morality. That is, does SSM predict that men and women will judge and/or behave differently in some systematic ways when it comes to fairness, honesty, generosity, risky heroism, and so on? Miller seems to think so. Indeed, he seems to think SSM would be supported over “sex-blind” accounts by any such sex differences: “sexual selection is probably relevant whenever there are cross-culturally stable sex differences in the display or judgement of human moral virtues” (2007: 102).

A prediction of subtle sexual dimorphism – unlike a prediction of stark sexual dimorphism – should not automatically cast SSM into doubt. Whether there are sex differences in human moral judgement and behaviour is an empirical issue that should not be prejudged. Before considering any empirical work, though, it is crucial to be clear about what findings would support SSM. Not just any sex differences will suffice to support SSM over “sex-blind” accounts of morality. To the extent that SSM does predict sex differences, it
predicts differences of a particular kind. As has perhaps been evident from my
discussion thus far, Miller’s presentations of SSM devote far more time to
discussing males’ moral displays than females’ displays. Miller even explicitly
states that, according to SSM, males are “predicted to allocate somewhat more
energy, time, and risk to mating effort, including costly, dangerous, public
displays of moral virtue” (2007: 102).

Miller does not make entirely clear whether he takes males’
supposedly greater inclination toward costly moral displays to be a standing
disposition (i.e. predicting that men more so than women will engage in such
displays across the board) or a conditional strategy (i.e. predicting that men
more so than women will engage in such displays when there is the chance of
attracting a mate). Miller (quite rightly) does not think costly moral displays
are always (or even often) consciously aimed at attracting mates (see §1.2.1),
but that is not to say such displays will be given willy-nilly.\textsuperscript{146} The question
for SSM must therefore be: is there evidence that men more so than women
engage in costly moral displays and, if so, under what circumstances?

Miller (2000) disappointingly relies largely on anecdote and appeal to
supposed common knowledge. “Waitresses”, he writes, “know more about
human generosity than most moral philosophers [and] report that groups of
men leave much better tips than groups of women, and men on dates with
women leave especially good tips” (2000: 325-326). However, SSM needs
data that has been collected under conditions more controlled than a
naturalistic restaurant setting (where selective flirting, for instance, may affect
tipping), and that has been subjected to statistical analysis more rigorous than
waitresses’ estimations.

Happily, Conlin \textit{et al.} (1999) have studied tipping. They found that
men tip better on smaller bills but women tip better on large ones: “percent-tip
is a decreasing function of bill size for both men and women [but] as the bill

\textsuperscript{146} Consider: the superb bird of paradise \textit{(Lophorina superba)} has nothing like human-level
cognitive sophistication and probably never thinks of his mating dance as a way to get sex, but
even so, he only ever busts his bizarre moves in the presence of a potential mate.
size increases, men’s percent-tips decline at a faster rate than those of women” (quoted in Andreoni & Vesterlund [2001: 306]). Men, it seems, tip more generously than women only when doing so is relatively cheap. This finding is at odds with what SSM would lead us to expect, and shows how reliance on anecdote and supposed common knowledge can lead one astray.

Miller (2007: 102) does somewhat better, citing studies that show men are over-represented in risky, prosocially-oriented professions like police work and fire fighting. As Miller himself admits, though, there are many factors operative here besides men’s efforts to signal their quality to potential mates. I believe that some of the empirical work done in experimental economics can help address our question about SSM and sexual dimorphism in morality in an appropriately rigorous way.

Experimental economics examines people’s tendencies to divide resources fairly or unfairly, generously or selfishly, to reward fairness and generosity, and to punish unfairness and selfishness, across a range of carefully-controlled conditions. One of the experimental economist’s basic tools is the so-called ‘money game.’ In the Dictator Game (DG), a subject (the dictator) must decide how to allocate a sum of money between themselves and a partner (the receiver), who has no option but to accept the dictated split, even if the dictator chooses to keep the whole sum. In the Ultimatum Game (UG), a subject (the proposer) must propose a way of dividing a sum of money with their partner (the responder), who can either accept the offer (in which case proposer and responder share the money as proposed) or reject the offer (in which case neither proposer nor responder gets anything). In the Public Goods Game (PGG), several subjects (the players) must individually decide whether to contribute their own resources to a common pool that will be multiplied by some factor then divided equally among all of the players. The problem each player faces is that the others might exploit them by hoarding their own resources while reaping the benefits of the player’s contribution. Often,

147 Davis & Holt (1992) provide an accessible introduction to the discipline.
complications are built into these basic games, as will be described in more detail below.

In what follows, I will survey a range of studies that investigated possible sex differences in money game behaviour. Throughout the following discussion, it should be kept in mind that subjects’ behaviour in the various money games is supposed to be indicative of how those subjects behave outside the limited context of the economist’s laboratory.\(^{148}\) The results are, to put it mildly, equivocal when it comes to supporting or undermining the prediction of SSM about sex differences in morality.

To begin with, I will present Andreoni & Vesterlund’s (2001) investigation of differences between men’s and women’s behaviour in the DG. The findings of this study match neatly with the pattern found by Conlin \textit{et al.} mentioned above. Andreoni & Vesterlund’s experimental design varied the amount that subjects playing the DG had to divide, with some subjects having a relatively large amount and others relatively little. Also varying across experimental conditions was the so-called “price of altruism” (2001: 294). To be clear, ‘altruism’ here simply measures how much the dictator gives the receiver: the more the receiver gets, the more altruistic the dictator. In one condition, dollars not claimed by the dictator were given to the receiver as per usual in the DG. In another pair of conditions, the exchange rate was such that dictators had to give up more than one dollar (either two or three) in order to give receivers a dollar (meaning, for example, that a dictator claiming half of the total to be divided would effectively be giving the receiver only a quarter of the total). In the final pair of conditions, for each dollar not claimed by the dictator, the receiver got more than one dollar (again, either two or three).

Andreoni & Vesterlund found that there was no simple answer to the question of which sex was more altruistic. Male dictators gave more to their partners than female ones when altruism was cheap, but female dictators gave more than male ones when giving a dollar required giving up two or three

\(^{148}\) See Binmore & Shaked (forthcoming) for criticism of this assumption.
dollars. In the experimental condition where altruism was most costly (with the smallest amount to divide and the worst exchange rate for giving), women were approximately twice as altruistic as men (2001: 298). Andreoni & Vesterlund thus concluded that “when it is relatively expensive to give, women are more generous than men; however, as the price of giving decreases, men begin to give more than women” (2001: 294). This finding is much like that of Conlin et al. reported above. It too conflicts with SSM’s prediction that men more so than women will be inclined toward costly generosity.

Sticking with the DG, we find Bolton & Katok (1995) asking whether there are sex differences in so-called “beneficent” behaviour, which in their study was measured by the amount of money the dictator allocated to the receiver. Bolton & Katok had subjects play numerous variations of the DG, varying the amount to be divided, the number of rounds of the game that were played, and the range of allocations that were possible. No sex differences in beneficent behaviour were found under any of the DG variations, leading Bolton & Katok to conclude that “there is no evidence of a gender effect on dictator-giving” (1995: 290).

Insofar as SSM predicts that men more so than women will be inclined toward costly acts of generosity and other morally laudable behaviours, we might expect that if SSM were right then Bolton & Katok would have found men being more generous in the role of dictator than women. Of course, it is open to a defender of SSM to say that the study lacked ecological validity to such an extent that it does not undermine SSM. I have quite some sympathy with this response. After all, the experimenters themselves say that “under typical experimental conditions, gender does not appear to influence the outcome [of the DG]” (1995: 291). “Typical experimental conditions” are those in which the proposer does not know the identity of the responder. The defender of SSM may insist that building in more detail to the experimental

---

149 On this last point, in some conditions dictators had to choose between splitting the money evenly and keeping everything, while in other conditions dictators had a wider range of options, such as an even split, keeping everything, or a 60/40 split (Bolton & Katok [1995: 289]).
conditions – in particular, letting money game players know the sex of their partners – would provide a better test of SSM’s prediction about sex differences.

As it happens, some such more richly-detailed studies have been carried out. Eckel & Grossman (2001) ran a study using the UG, asking whether there are sex differences in UG play when players know the sex of their partner. It is worth making explicit the complexity of this question. The comparisons of interest are between: men’s offers to men; men’s acceptance of offers from men; men’s offers to women; men’s acceptance of offers from women; women’s offers to women; women’s acceptance of offers from women; women’s offers to men; women’s acceptance of offers from men.

Eckel & Grossman lay out three hypotheses to be tested (2001: 174). The “compete/cooperate” hypothesis predicts that men will make less generous offers than women and women will be more accepting of low offers than men. The “chivalry” hypothesis predicts that men will make more generous offers when paired with a female responder than with a male responder and men will be more accepting of offers from female proposers than offers from male proposers. The “solidarity” hypothesis predicts that proposers will be more generous toward responders of the same sex and responders will be more accepting of offers from proposers of the same sex. Of these three, SSM seems most akin to the chivalry hypothesis. After all, if sexual selection has favoured costly displays of generosity as signals of mate quality, and has favoured such displays even more strongly in men than in women, then the very highest levels of generosity should be seen in men with the chance to act generously toward women (either by offering more or accepting less).

Eckel & Grossman found that women were, overall, slightly more generous proposers than men and were also, overall, more accepting of low offers, thus supporting compete/cooperate (2001: 181). In addition, solidarity was supported for female pairs (2001: 183). Men, though, did not propose significantly more generous divisions to women than to other men. While
Eckel & Grossman take the fact that any given offer was more likely to be accepted if it came from a woman than if it came from a man to provide partial support for chivalry, to me this is not clear. It is true that men were more likely to accept low offers from women than from men, but so were women (since female/female pairings showed solidarity). It is not clear from the data that men were even more accepting of low offers from women than were women. Without that result, taking the data to support (even partially) chivalry seems to me rather strange. Finally, it was not the case that the most generous proposers were men making proposals to female responders (Eckel & Grossman [1995: 177]). In sum, while Eckel & Grossman’s experimental design seemed conducive to revealing sex differences relevant to SSM, the data did not fit with what we can reasonably take to be the predictions of SSM.

Moving on now from the DG and UG to the PGG, we again find mixed results as far as SSM is concerned. Brown-Kruse & Hummels (1993) investigated sex differences in contributions to a public good. In their experiment, all-male and all-female groups played a PGG. They found that “males contributed at higher rates than did women” (1993: 264). If we see contributing to the public good despite the risk of being exploited by free-riders as a kind of costly altruism, then Brown-Kruse & Hummels’ finding accords well with SSM’s prediction. However, a similar study by Nowell & Tinkler (1994) reported a conflicting result. Nowell & Tinkler had all-male, all-female and mixed groups play a PGG. They found that “all-female groups tended to make higher contributions than all-male or mixed gender groups” (1994: 29).

The conflicting results here may be due to a difference in experimental design across the two studies. In the first study, subjects had to choose between contributing and not contributing their $1 to the public good. In the second study, subjects had to choose how many (if any) of their 62 tokens to contribute to the public good. Contribution to the public good was thus an all-or-nothing affair in the first study, while in the second study subjects could hedge their bets, so to speak. This is, I think, an important difference between the studies that makes comparing their findings problematic. In any case, the
results of these studies are equivocal as far as SSM’s predictions about sex difference are concerned.

Having surveyed several studies using the various money games described earlier in this section, I will close by discussing some work in experimental economics that does not use money games but does seek to uncover sex differences that seem relevant to SSM.

Eckel & Grossman (1996) investigated sex differences in decisions involving fairness, again with mixed results as far as SSM is concerned. Subjects in their study had to choose between equally sharing a relatively large sum with an unfair player or a relatively small sum with a fair player. In the “low cost” condition, the larger sum was $12 and the smaller sum was $10. In the “high cost” condition, the larger sum was $12 and the smaller sum was $8.

Eckel & Grossman view opting to share the smaller amount with the fair player as an indication of “willingness to reward fairness” (1996: 193). There is, however, another way to see things here. Opting not to share the larger amount with the unfair player may be interpreted as paying a cost to punish unfairness. By choosing to share with the fair player, the subject ensures the unfair player gets nothing. Sharing the smaller amount with the fair player can, I think, be seen as a kind of costly, moralistic punishment of the unfair player. As such, SSM would lead us to expect that men more so than women would opt for the smaller amount.

While Eckel & Grossman found some sex differences in subjects’ decisions, the upshot of their study for SSM is mixed. In the high cost condition, men were significantly more likely than women to choose to share the smaller amount with the fair player (1996: 152). That men more so than women were willing to punish unfairness when doing so was relatively costly jibes well with SSM’s prediction that men more so than women will engage in

---

150 Both players were in fact fictional. Subjects were just led to believe that the unfair player had, earlier in the experiment, dictated an $18/$2 split to a partner in the DG, and were told that the fair player had dictated a $10/$10 split.

151 40% of men compared to 33% of women took this option.
costly displays of moral quality. However, Eckel & Grossman also found that overall (i.e. pooling across both conditions), men were significantly less likely than women to punish unfairness (1996: 153). This finding seems less in line with the predictions of SSM.

The last study I will discuss here is that by Iredale et al. (2008). This study sought to discover how being observed by members of the same or the opposite sex affects charitableness in men compared to women. Subjects in this study were seated at a computer with either a member of the same sex, the opposite sex, or alone. Subjects first played a computer game in which they could earn real money and were then given an on-screen option to donate some of their earnings to charity. After leaving the computer, subjects were asked to rate the attractiveness of their opposite-sex observer, if they had had one.

Iredale et al. predicted that men under observation by women that they (i.e. the men themselves) considered attractive would be more charitable than unobserved men but that men under observation by women they considered unattractive, as well as men under observation by other men, would not be more generous than unobserved men. SSM can reasonably be taken to make the same prediction, for the reason laid out when discussing Eckel & Grossman's study. Happily for SSM, Iredale et al. found that men were more charitable when observed by an attractive woman but not when the observer was another man or (according to the subject) an unattractive woman (2008: 389). Women's levels of charity, by contrast, did not significantly vary from the baseline set by subjects in the unobserved condition, regardless of whether their observer was an attractive or an unattractive male or another female (2008: 389). Iredale et al. take these findings as evidence that conspicuous male charity may function as a mating tactic (2008: 391). Here, at last, is some evidence congruent with the predictions of SSM.

---

152 The observers were actually collaborators in the experiment.
153 Data from self-reported homosexual subjects was excluded from the final analysis.
154 Building upon the work by Iredale et al. in a way that accommodates the possibility of intra-sexual competition and signalling would be a worthwhile project. For example, it would be interesting to see the effect on men’s donations of being observed by an attractive female
In light of the survey above, it seems Miller has painted an overly simplistic picture of sex differences in charitable donation. The reality is far more complex and not nearly so flattering to SSM. Of course, I do not mean to suggest the verdict is in and that SSM has been ruled against. As was mentioned at the outset of the survey, there is uncertainty inherent in moving from observations made under artificially sparse experimental conditions to claims about how people behave during complex social interactions like those involved in inter-sexual signalling and mate choice. Also, as has been stressed throughout my discussion, the findings on sex differences in generosity, fairness, and so on are currently inconclusive. The verdict is simply not yet in on sex differences in morality. My aim in this section has been simply to emphasise that the support for SSM’s predictions about subtle sexual dimorphism in human morality may well be less sturdy than Miller (along with common knowledge and anecdote) would have it.

5.3. Assessing the Evidence for SSM

In Chapter 1 (see §1.2.2d), I considered the empirical evidence presented by Miller (2000) for SSM. I focused there on Miller’s discussions of Hawkes’ (1991) show-off hypothesis and Buss’ (1989) mate preference survey. I highlighted the shortcomings of these as evidence for SSM. Recall, the ‘show-off’ hypothesis faces serious empirical difficulties, while the findings of Buss’ survey cannot distinguish between the claim that moral behaviour is an inter-sexually selected signal and the claim that such behaviour is merely a cue used in mate choice. My discussion in Chapter 1 closed by laying out what must be done in order to empirically substantiate a costly signalling claim about moral behaviour (see §1.2.2d). Recall, to empirically substantiate a costly signalling claim about moral behaviour, the following questions must be answered in the affirmative:

(1) Do potential mates respond differentially to displays of moral behaviour?
(2) Does moral behaviour indicate qualities of interest to potential mates?

and several other men. Perhaps the presence of potential sexual rivals would prompt even greater generosity.
(3) Is moral behaviour a reliable indicator of such qualities?
   (a) Is moral behaviour costly in terms of some component of fitness?
   (b) Is moral behaviour differentially costly for honest versus dishonest senders?

It must be admitted at the outset that the currently available empirical evidence does not even come close to substantiating SSM. Studies measuring the fitness costs and benefits of moral behaviour in the context of mate choice are, regrettably, unavailable. The best I can hope to do in this section is to amass findings that lend a degree of indirect support to the idea that moral behaviour serves as a sexually-selected costly signal. Below, I will present some empirical support for the claim that paradigmatically moral behaviours – such as being generous, fair, trustworthy, and punishing those who are not – play a signalling role. This is, as far as it goes, good news for SSM. However, as will be detailed later (see §5.4), the very evidence supporting the claim that moral behaviour can serve as a signal creates a problem for SSM. That evidence suggests that moral behaviours play a signalling role in contexts beyond mate choice, which in turn suggests SSM must be only part of a more general signalling-based explanation for morality. For now, though, I will focus on the evidence supporting a signalling role for moral behaviour.

5.3.1. A case study: Meriam turtle hunting

I will begin by revisiting the show-off hypothesis. Recall from Chapter 1 (see 1.2.1) that Miller takes Hawkes’ work to provide a possibility proof of sorts for SSM, by showing that an appeal to costly signalling and sexual selection to account for puzzlingly costly, other-benefiting behaviours is a viable explanatory strategy. Of course, this alone does not do a great deal to empirically substantiate SSM. Appealing to costly signalling and sexual selection could be a viable explanatory strategy in the case of big game hunting but, in the end, quite hopeless when it comes to morality. Even so, putting such a thing in place is desirable, if it can be done. In this section, by presenting some recent anthropological research, I will show that it can be. As well as providing indirect support for SSM, this research also holds tantalising
hints of more direct support for the claim that conspicuous moral displays can serve as costly signals and pay off reproductively, as I will detail below.

Anthropologist Rebecca Bliege-Bird and her colleagues have studied patterns of food sharing among the Meriam islanders of Australia’s Torres Strait. Bliege-Bird and colleagues were interested in whether Meriam food sharing could be explained by the “risk reduction reciprocity” (RRR) hypothesis. RRR is essentially the same as what Gurven (2004: 545) calls the “reciprocal altruism” (RA) hypothesis (see §1.2.2d). RRR states that “cooperating foragers pool their harvest after foraging, and then consume more or less equal shares from this common pot, reducing their own variance in consumption” (2002: 298). Bliege-Bird and colleagues argue that RRR cannot explain food sharing amongst the Meriam when the food in question is turtle meat.

Meriam islanders obtain turtles in two kinds of ways. Turtle collecting is a matter of flipping and trussing turtles as they crawl onto the beach to lay their eggs during the nesting season, and is a fairly sedate activity engaged in by young and old men and women. Turtle hunting, by contrast, is far more strenuous. It requires going to sea to capture turtles swimming near reefs, and is done only by men. A turtle hunting party is composed of a leader (who organises and directs the hunt), a tillerman (who steers the boat) and a jumper. The jumper’s job it is to tie a rope around himself then leap out of the boat onto the turtle and grapple with it while the other men use the rope to haul both jumper and turtle aboard (Bliege-Bird et al. [2001: 11]).

Bliege-Bird et al. (2002) analysed patterns of turtle-meat sharing among the Meriam. They found no evidence of contingency in the sharing of turtle meat. Households that provided a lot of turtle meat to others did not receive more turtle meat than households that provided less. Nor were households that hoarded their own turtle-meat excluded from sharing in other households’ hauls (i.e. free riders were not discriminated against). Bliege-Bird et al. thus concluded that “Meriam food sharing is not structured in the conditional manner required to stabilise RRR” (2002: 315).
Further work by Bliege-Bird and colleagues considered an alternative to RRR as an explanation for turtle meat sharing, namely, costly signalling:

[T]he benefits of [turtle] hunting are gained through the honest-signal value of acquiring the prey, rather than through consumption, and that honesty is maintained through differential costs and benefits: men of higher phenotypic quality benefit more or pay less than lower-quality individuals per unit signal (Bliege-Bird et al. [2001:14]).

Bliege-Bird et al. focus on turtle hunting in presenting their costly signalling hypothesis, as opposed to turtle collecting, for two reasons. First, as mentioned above, collecting is sedate whereas hunting is strenuous. Second, while collected turtle meat tends to be shared privately among small numbers of households, hunted turtles are exclusively used to provision large public feasts. These elaborate events can bring together several different villages, and the identities of the providing hunters become common knowledge. Compared to collecting, then, hunting seems both more costly and more apt to serve as a signal.

Bliege-Bird and colleagues note that turtle hunting as a signal has good “broadcast efficiency,” which is a matter of “the number of observers attracted per unit of signalling effort” (Smith & Bliege-Bird [2006: 123]). The reason, they say, is that:

...we can expect that responders will prefer signals that provide a collective good worth $G$ over some equally informative signal that provides no collective good because, in addition to the gains from the information transferred in the signal, each of the $n$ responders’ payoffs will also be increased by $G/n$ (2006: 125).

As a brief aside, it is worth mentioning that the notion of broadcast efficiency can supplement Miller’s account of why costly signals of mate quality tend to be other-benefiting rather than merely wasteful. Miller suggests that group selection acting on cultural groups playing different “courtship games” – some that generate benefits for the group members and some that are purely wasteful – will lead to the predominance of groups whose members signal to potential mates in other-benefiting ways (2000: 314-318). I find Miller’s suggestion quite plausible. In mentioning the notion of broadcast efficiency, I
merely mean to note a way in which those who signal their quality via other-benefiting acts may also benefit themselves: more people will pay attention to their signals.

Returning now to the case study at hand, Bliege-Bird and colleagues begin their case for a costly signalling account of turtle hunting by pointing out the various costs of turtle hunting. It is worth pointing out the need to keep in mind here the clarificatory work done in Chapter 4 (see §4.2.1). What Bliege-Bird and colleagues really do is identify the expenditures required by turtle hunting. They do not demonstrate that turtle hunting negatively impacts on some component of hunters’ fitness. The inference from expenditures to costs remains to be made. Even so, the expenditures Bliege-Bird and colleagues identify do much to make this inference plausible.

Turtle hunts take time to organise and conduct. They require resources: boats, nets and money for fuel. They are energetically demanding and pose a risk of injury or even death for all hunters but especially for the jumpers. Turtle hunts are also more abstractly perilous, since failing to deliver a turtle for a feast has significant social costs: “everyone will know when a hunter has failed” (Bliege-Bird et al. [2002: 17]). Taken in conjunction, the considerable expenditures of time, resources, and effort required by turtle hunting, and the physical and social risks it involves, strongly suggest that turtle hunting should be counted as a costly activity.

Bliege-Bird and colleagues continue their case for a costly signalling account of turtle hunting by identifying the qualities plausibly signalled by successful turtle hunting (Bliege-Bird et al. [2001: 15]; see also Smith et al. [2003: 124]). A distinction should be drawn, they suggest, between jumpers and hunt leaders when considering the qualities successful turtle hunting signals. Successful jumpers plausibly advertise their physical qualities like strength, stamina, agility, and perception. Successful hunt leaders plausibly advertise their cognitive and leadership abilities as well as their prosocial tendencies. A successful turtle hunt shows that the hunt leader knows about turtles and the sea, can effectively organise a group of men, and is willing to
devote himself to providing an important public good.\textsuperscript{155} Bliege-Bird \textit{et al.} have only plausibility arguments to offer regarding the connection between success at turtle hunting and these kinds of underlying qualities. Even so, their plausibility arguments are quite plausible.

Showing that turtle hunting is differentially costly for hunters that have the relevant qualities and those that do not is, Bliege-Bird \textit{et al.} say, difficult, since “low-quality individuals may face a large enough risk of failing [that they expect] a net social deficit from their signalling attempts, and hence may avoid signalling at all” (2001: 17). Bliege-Bird \textit{et al.} offer a rather circuitous argument to support the idea that turtle hunting is differentially costly. The key claims are as follows. Older hunt leaders are typically more skillful than younger hunt leaders. Nesting season hunts are easier than hunts outside nesting season. Outside the nesting season, the only way to obtain a turtle is by hunting. The signalling value of a turtle is thus highest outside of nesting season, due to the difficulty of hunts during that time and the lack of scope for faking the signal (by passing a collected turtle off as a hunted one). Bliege-Bird \textit{et al.} argue that if turtle hunting were \textit{not} differentially costly, then high- and low-quality hunters alike would operate outside of the nesting season. In fact, hunt leaders operating during the nesting season are on average younger than hunt leaders operating outside the nesting season (2001: 15). Bliege-Bird \textit{et al.} take this as support, albeit quite roundabout, for the claim that hunting is

\textsuperscript{155} There is a complication that bears mentioning here. Bliege-Bird \textit{et al.} report that the leader occupies the high-prestige role among turtle hunters:

Hunt leaders are invested with public recognition and receive full credit for the kill regardless of whether or not they directly participate in capture... Jumpers are rarely individually credited by others for acquiring a turtle, and drivers are rarely publicly credited with participation in the hunt. Feast-goers can readily name hunt leaders as providers of the turtle, even several years after the feast, but when pressed for the names of other hunt participants, lump all others together as “the boys” (2001: 11, 15).

In light of this, one might wonder why jumpers, who play the most obviously dangerous role in the hunt, bother. Bliege-Bird \textit{et al.} provide the answer:

Jumpers may begin in this role as young as 15–17 years old; anecdotal evidence indicates that those who excel in this role and gain increasing knowledge and peer respect eventually become hunt leaders, while others remain jumpers or cease participating in turtle hunting (2001: 15).

The best jumpers, it seems, eventually move into the high-prestige leader role.
differentially costly in the kind of way required for a costly signalling account of the practice to work.\footnote{Bliege-Bird and colleagues' argument can be supplemented with some fairly uncontroversial comments. First, if one of the costs of hunting is the fuel needed for the hunting boat, then less knowledgable hunt leaders will most likely incur greater costs in this respect than more knowledgable leaders, as they will most likely spend longer looking for turtles. In the same vein, if another of the costs of hunting is the risk involved, then spending more time on a hunt should (all else being equal) equate to a greater chance that something will go wrong during hunts run by less knowledgable leaders. The case for differential costs when the costs are risks is even clearer for jumpers. A weaker, less robust jumper surely runs a greater risk of injury than a stronger, more robust one, making turtle hunting differentially costly across such individuals. Lastly, consider the social costs of hunting, that is, the risk of embarrassment and loss of status should the hunter fail to provide a turtle for a public feast. This risk will clearly be higher for less competent hunters.}{156}

The final element in the case for turtle hunting as costly signalling is work by Smith et al. (2003) estimating the reproductive success (RS) of turtle hunters and their mates.\footnote{For details on exactly how individuals' RS was quantified, see Smith et al. (2003: 117).}{157} Smith et al. report that turtle hunters begin reproducing earlier than non-hunters, have more mates than nonhunters, have higher quality mates than nonhunters, and are more able than nonhunters to attract young sexual partners later in life.\footnote{Mate quality here is reckoned by the Meriam themselves: "a measure of mate quality that is culturally recognised by Meriam is whether one's spouse is "hard working" [and] hunter's mates received significantly more nominations as hard working than nonhunter's mates" (Smith et al. [2003: 120]).}{158} Taken together, these factors result in turtle hunters' RS being approximately two and half times greater than that of nonhunters (Smith et al. [2003: 118]). Smith et al. also found that the mates of turtle hunters had RS significantly greater than that of other Meriam women (2003: 123), suggesting that women benefit reproductively from mating with turtle hunters.

Recall again the empirical burden on a costly signalling claim about some behaviour X:

\begin{enumerate}
\item receivers respond differentially to X
\item X is an indicator of qualities of interest to receivers
\item X is a reliable indicator of such qualities
\begin{enumerate}
\item X is costly in terms of some component of fitness
\item X is differentially costly for honest vs. dishonest senders
\end{enumerate}
\end{enumerate}

The point at which Bliege-Bird and colleagues' costly signalling account of turtle hunting is weakest is (1). Bliege-Bird and her collaborators

156 Bliege-Bird and colleagues' argument can be supplemented with some fairly uncontroversial comments. First, if one of the costs of hunting is the fuel needed for the hunting boat, then less knowledgable hunt leaders will most likely incur greater costs in this respect than more knowledgable leaders, as they will most likely spend longer looking for turtles. In the same vein, if another of the costs of hunting is the risk involved, then spending more time on a hunt should (all else being equal) equate to a greater chance that something will go wrong during hunts run by less knowledgable leaders. The case for differential costs when the costs are risks is even clearer for jumpers. A weaker, less robust jumper surely runs a greater risk of injury than a stronger, more robust one, making turtle hunting differentially costly across such individuals. Lastly, consider the social costs of hunting, that is, the risk of embarrassment and loss of status should the hunter fail to provide a turtle for a public feast. This risk will clearly be higher for less competent hunters.

157 For details on exactly how individuals' RS was quantified, see Smith et al. (2003: 117).

158 Mate quality here is reckoned by the Meriam themselves: "a measure of mate quality that is culturally recognised by Meriam is whether one's spouse is "hard working" [and] hunter's mates received significantly more nominations as hard working than nonhunter's mates" (Smith et al. [2003: 120]).
currently cannot (as they acknowledge) rule out the possibility that hunting success and reproductive success correlate, not because the former facilitates the latter, but because the two have a common cause: perhaps healthier, stronger, more capable individuals are more likely to do well at both. Recognising this fact, Bliege-Bird and colleagues write:

The key issue is whether the underlying quality (e.g., greater health and vigour) is observable by simple visual inspection or whether it is difficult to directly observe but can be reliably signalled by activities such as success in turtle hunting (Smith et al. [2003: 122]).

That is to say, a costly signalling account of turtle hunting is compatible with some quality or qualities causally contributing to both hunting and reproductive success, so long as hunting prowess figures somehow in Meriam folks’ decision-making. That, however, remains to be shown. Meriam women do not in fact express an explicit preference for turtle hunting prowess in a mate, which leads Bliege-Bird and colleagues to speculate that hunting prowess may be a signal in intra-sexual competition among males for social status (Smith et al. [2003: 125]). At this point, though, speculation is all that is on offer.

To sum up, turtle hunting by Meriam islanders provides a good candidate for a case of signalling via costly and conspicuous other-benefiting behaviour, even if the evidence is currently incomplete. This case study supports SSM indirectly, by making it plausible that costly signalling theory and the theory of sexual selection can explain behaviour that is, like much paradigmatically moral behaviour, costly and other-benefiting.

I will conclude this section by mentioning a tantalising hint, drawn from Bliege-Bird and colleagues’ ethnographic observations, that food sharing among the Meriam may be directly relevant to claims about the signalling value of moral behaviour. Bliege-Bird et al. observe that:

159 Consider, by way of parallel, the peacock’s tail. A high-quality peacock’s ‘good genes’ are what allow him to produce his vibrant tail and are also what provide his offspring with a better-than-average chance of survival and thus increases his fitness. That the same quality is causally important in both cases is not undermining of the costly signaling claim here. The crucial thing is that the tail itself has some causal impact on peahens’ behavior.
...[a] common Meriam ideal of generosity involves reference to the concept of *debe tonar*, which means "the good way." *Debe tonar* constitutes a set of principles to guide everyday social interactions, and adhering to those principles is considered a signal of Meriam moral and ethnic identity. Following *debe tonar* is said to provide long term benefits in the form of an enhanced social reputation as a "good person" (2002: 316).

The work by Bliege-Bird and her collaborators thus does far more for SSM than Hawkes’ showoff hypothesis, because it explicitly links food sharing with morality. Miller at times (e.g. [2008: 216]) tries to claim Hawkes’ work as direct support for SSM:

[I argue] that many human virtues evolved through sexual selection as costly signals, as fitness indicators. This hypothesis has been advanced by a few previous researchers, e.g. Hawkes (1991).

The problem, of course, is that Hawkes advanced a hypothesis concerning hunting and food sharing, not moral virtue. As far as the showoff hypothesis was concerned, big game hunters could simply be self-aggrandising thrillseekers whose toughness is warily respected by the recipients of their bounty. In contrast, Meriam hunters who provide turtles for feasts conspicuously display their adherence to the moral values prized in their culture. They are recognised not just as tough guys but as good people, and are rewarded as such. This is indeed encouraging evidence for SSM.

5.3.2. Competitive Altruism

I have already in this chapter introduced the basic methodology employed in experimental economics (see §5.2.2) and discussed some studies relevant to SSM. In this section, I will return to that field. Experimental economists abstract away from much of the messy real-world detail that makes life difficult for anthropologists. Of course, abstraction brings its own problems: experimental simplicity may come at the cost ecological validity. Ideally, though, careful study design allows for powerful tests of hypotheses about human economic behaviour. Whether this ideal has been achieved must be assessed on a case-by-case basis.
In what follows, I will survey a number of experimental economics studies. To appreciate the relevance of these studies to SSM, recall once more the empirical burden on a costly signalling claim about a trait X:

(1) receivers respond differentially to X
(2) X is an indicator of qualities of interest to receivers
(3) X is a reliable indicator of such qualities
   (a) X is costly in terms of some component of fitness
   (b) X is differentially costly for honest vs. dishonest senders

The results below provide support, I claim, for (1) and (2), when the trait in question is behaving fairly and generously. Many of the studies I will discuss were anticipated (and perhaps inspired) by a paper of Gilbert Roberts’ (1998). Roberts noted that:

...[m]any interactions between organisms occur within partnerships, yet partners change... [A] realistic way in which to model interactions may be to have a two stage model incorporating both an ‘assessment’ stage and a ‘partnered’ stage. In the assessment stage, animals interact with the whole population, whereas in the subsequent paired stage they interact solely with their chosen partner (1998: 428).

Roberts went on to present what he called the “competitive altruism” hypothesis, according to which:

...individual differences in generosity combined with partner choice can lead to competition for partners and thereby to an escalation in generosity which is rewarded by greater benefits for those which choose to or can invest more (1998: 428).

Many of the studies below use Roberts’ two-stage design and their results make clear that Roberts’ hypothesis was on target. The first study I will present here was conducted by Pat Barclay (2004), who investigated the effect of opportunities for reputation-building on altruism and the effect of altruism on individuals’ judgements of others’ trustworthiness. Barclay first had groups of four subjects play a five-round PGG. ‘Altruism’ in this study was a matter of how much a subject contributed to the public good: the more a subject contributed, the more altruistic that subject. Subjects’ anonymity was

---

160 ‘Altruism’ thus operationalised makes no reference to the psychology of the giver. It is compatible with the results of this study (and those that build on it, to be described below) that highly ‘altruistic’ individuals were motivated entirely by self-interest, in particular a prudent
preserved by pseudonyms, but each player knew how much each other pseudonymously-named player had contributed to the public good. Barclay then had subjects play a trust game, which worked as follows: each player was given a sum of money and the option of sending some or all of that money to another player or players (the trustees). Any money sent to a trustee would be tripled, and the trustee would then have the option of returning some, all or none of that amount to the sender. In one condition, subjects were unaware that a trust game would follow the PGG (call this the ‘no reputation building’ condition). In the other conditions (the ‘reputation building’ condition), subjects knew the trust game was coming. In one of the reputation building conditions, subjects could opt to send money to up to three of the other players during the trust game (the ‘non-competitive’ condition). In the other reputation building condition (the ‘competitive’ condition), subjects could entrust money to only one other player.

Barclay found that subjects were significantly more altruistic during the PGG in the reputation building conditions than in the other condition, and contributions to the public good were highest in the competitive reputation building condition (2004: 214). In the non-competitive reputation building condition, the highest contributor to the public good was entrusted with the most money by other group members (2004: 215). In the competitive reputation building condition, there was, interestingly, no significant difference in the amount of money entrusted to the three highest-contributing players: the lowest-contributing player was, however, entrusted with nothing (2004: 216).

Barclay’s results support the claim that risky generosity affects others’ behaviour toward an agent. More altruistic individuals were entrusted with more money and chosen more often as trust-game partners. Barclay’s results concern for their reputation. It is also possible that subjects who made large contributions to the public good did so in a genuinely cooperative spirit. The point of these studies as far as SSM goes, however, is not so much what they show about signallers but rather about receivers. The interesting finding, as will become apparent in the discussion above, is that subjects respond to ‘altruistic’ acts by making inferences about the character of the ‘altruist.’ Insofar as subjects bring their everyday habits of thought into the laboratory, this is an important finding.
also support the claim that risky generosity is taken to indicate a further quality of interest by other agents. Generous contributors to the public good were seen as trustworthy as a result of their generosity.

A pressing question, insofar as SSM is concerned, is whether the trust placed in altruists was warranted: did more altruistic individuals return more money to their trustees than less altruistic ones? Barclay reports that “the only significant predictor of the proportion of money individuals were willing to return in the trust game was the amount that the individuals themselves sent to other players” (2004: 218). That is, more altruistic subjects did not return significantly more entrusted money than less altruistic subjects.

Barclay’s finding might be explained in a way consistent with SSM by noting that, in Barclay’s experimental set-up, there was strong incentive for – and few barriers to – deception, especially in the competitive reputation building condition. Indeed, subjects seemed cognizant of this fact. As mentioned above, in the competitive reputation building condition the lowest PGG contributor was not entrusted with any money, but there were no significant differences in the amounts entrusted to the other players. This finding makes sense if subjects were sceptical of altruism as a signal when the incentive for deception was highest. Perhaps subjects avoided the clearly bad bet (the lowest PGG contributor) and were then indifferent between the remaining options.

Pat Barclay & Robb Willer (2007) built on Barclay’s (2004) study of altruism. Barclay & Willer’s subjects were placed into groups of three, with members designated A, B, and C. First, A and B played a game in which each was given a sum of money and could choose to send some, all or none to the other player. Any money sent was doubled before being given to the other player, and each player had to decide how much (if any) to send before finding out what the other player had decided. Next, C was paired with either A or B. In one condition, C was then randomly paired with A or B and not told about their partner’s behaviour in the first game. In another condition, C was then randomly paired with A or B and was informed of their partner’s contributions
in the first game. In the final condition, C was then told about both A and B’s behaviour in the first game and allowed to choose either A or B as a partner. Subject C and their partner then played the same game that A and B played initially. In each condition, A and B knew how the pairing for the next round of the game would be determined.

Barclay & Willer found that A and B sent more money to each other in the second and third conditions than in the first condition (2007: 751). They also found that generosity influenced the behaviour of C, who “almost always chose to interact with the more generous member of a pair” (2007: 752). Finally, it appeared that generosity influenced C’s behaviour differently depending on whether or not A and B were competing to be chosen as partners:

[W]hen participants could view their partners’ past donations, they gave significantly more to high contributors than to low contributors when those partners were assigned randomly... but not when they got to choose partners (Barclay & Willer [2007: 751]).

These findings match with those of Barclay (2004). Going beyond Barclay (2004), however, Barclay & Willer found that:

...participants’ earlier donations were correlated with their later donations when there was no competition to be chosen... but were not significantly correlated with their later decisions when they had been competing to be chosen (2007: 751).

That is, initial generosity was a reliable indicator of future generosity when reputation building was not competitive, but when reputation building was competitive, deceptive signalling of generosity increased.

The upshot of the two studies just presented is that competition for inclusion in profitable partnerships can prompt individuals to be more generous. It seems, though, that the very fact of competition undermines the value of generosity as a signal, at least when deception is cheap. I’ll now present some research that shows that the cost of signalling influences people’s judgement of the reliability of the signal.
Rob Nelissen investigated "how the costs invested in an altruistic act influence its interpersonal consequences" (2008: 243). By 'altruism,' Nelissen meant moralistic punishment, specifically, the paying of a cost to punish unfairness. He predicted that "people [would] confer social benefits (both in terms of enhanced preference and financial rewards) on altruistic punishers proportionally to the cost they incurred in punishing" (2008: 243-244).

Nelissen’s subjects were given a sum of money with which to play a trust game (as described above) with a partner drawn from among the participants of a previous experiment. Subjects were told that their three potential partners – labelled A, B, and C – had observed a DG in which the dictator split $10 unevenly with the receiver, keeping $8 and giving only $2. Subjects were also told that A, B, and C had had the opportunity to spend some of their own money to take money away from the dictator: giving up $1 would reduce the dictator’s total by $2. Finally, subjects were told that A chose to spend $0 out of $5 on punishment, B chose to spend $1.50 out of $5 on punishment, and C chose to spend $1.50 out of $10 on punishment. In one condition, subjects were randomly matched with A, B, or C and were then asked how much they would entrust to that partner. In the other condition, subjects were asked which of A, B, and C they wanted to play the trust game with and were then asked how much they would entrust to that partner.

Nelissen found that subjects with a choice chose B over A and C (2008: 244). He also found that, when pairing was random, subjects paired with B sent the most money in the trust game (2008: 246). As Nelissen interprets the findings:

[T]he costs incurred in altruistic punishment were perceived as signalling the extent to which punishers value fairness... [P]eople prefer punishers more [as trust-game partners] if they invest more to punish unfairness but only if the invested amount can be perceived as a reliable index of fairness concerns (2008: 244, 246).

The difference between B and C lies in the relative cost each paid to punish unfairness. Punishment was, relatively speaking, twice as costly for B

---

161 The 'partner' was in fact fictitious.
as for C. Subjects thus seem to take the cost of punishment into account when deciding who to interact with or how to behave toward partners that are forced upon them. While this does not show that more costly acts of punishment are more reliable signals that the punisher values fairness, it at least suggests that observers judge them to be such.

Nelissen’s work on moralistic punishment dovetails with Barclay’s (2006) work on the topic. Barclay ran an experiment in which accepting a cost to punish free-riding during a PGG benefitted other players. He found that individuals who paid to punish were subsequently rated as more trustworthy and more worthy of respect than non-punishers and were chosen over non-punishers as partners in subsequent trust games (2006: 330).

To sum up, the experimental economics research presented above supports the idea that costly, other-benefiting acts – being generous, fair, and paying to punish those who are not – can induce observers to trust the agent, to preferentially include the agent in profitable interactions, and to treat the agent generously and fairly in turn. The above research also shows that observers are sensitive to costliness when judging the reliability of such acts as indicators of character or future behaviour (placing the most faith in the most costly displays).

What findings from experimental economics do not do very convincingly is demonstrate how costs can maintain the reliability of other-benefitting behaviours as signals. This may well be because the costs of deception in a laboratory setting are relatively small, given that the amounts of money in play are typically small to negligible. Even so, as noted by Barclay & Willer, “participants tend to bring their expectations and preferences from the outside world into experiments” (2007: 752). Subjects’ money game play may thus reflect their behaviour in real-world situations, where deceptive

---

162 The free-rider was (unbeknownst to subjects) a computer player that responded to punishment by increasing its contributions to the public good in subsequent rounds of the game.
signalling of generosity, fairness, and moral quality more generally is (for the reasons covered in §4.3.3) genuinely costly.

Recall once more the empirical burden on a costly signalling claim:

(1) receivers respond differentially to X
(2) X is an indicator of qualities of interest to receivers
(3) X is a reliable indicator of such qualities
   (a) X is costly in terms of some component of fitness
   (b) X is differentially costly for honest vs. dishonest senders

The results from experimental economics discussed above provide support for (1) and (2), when the behaviours in question are fair and generous ones. As is probably apparent, though, fellow players in the money games described above are not potential mates. Rather, they are potential partners in economic interactions. This observation leads us directly into the next section of this chapter.

5.4. The Many Receivers Problem

In the previous section, I presented empirical work that appeared to provide some support for SSM. The work I presented does support a costly signalling account of moral behaviour but is not entirely good news for SSM. As I have already intimated, it can serve as the basis for a criticism of SSM. The criticism is that SSM is too narrowly focused. The empirical work presented above indicates that, if a costly signalling story is the right kind of story to tell about moral behaviour, then the story is far more complex than SSM, involving more receivers than just potential mates. I will call this the ‘many receivers’ problem for SSM.

Consider again the anthropological research into Meriam turtle hunting conducted by Bliege-Bird and her colleagues, who suggest that:

...turtle hunting provides evidence of skills that could honestly signal several relevant qualities: environmental and ethological knowledge, risk-taking, strength and agility, leadership and organisational abilities, and prosocial tendencies e.g., expending time and energy in providing goods for public consumption at feasts (Bliege-Bird et al. [2001: 15]).
It is not only potential mates to whom these qualities would be of interest. In fact, as mentioned above, Bliege-Bird and her colleagues believe that the reproductive benefit of turtle hunting does not come via mate choice (i.e. inter-sexual selection). Rather, they think turtle hunting serves its signalling function in the context of intra-sexual selection:

[O]ur view [is] that men gain access to high-quality mates by status competition with other men, rather than by using hunting to signal directly to potential mates... We believe that the primary reasons better hunters fare well in mate choice is that they gain higher status and social dominance, which in turn are cues used by women in mate choice (Smith et al. [2003: 123, 125]).

To be fair, Miller is cognizant of the possibility that intra-sexual selection as well as inter-sexual selection may be important in understanding the evolution of morality. He distinguishes between “direct mate choice for moralistic displays during courtship” and “the indirect reproductive benefits of high status” and claims that “both effects were probably important during human evolution” (2000: 307). Miller, I imagine, would be happy to agree that some moral behaviours may serve as signals within the context of intra-sexual rather than inter-sexual competition.

The many receivers problem cannot be so easily squared with SSM, though. Sexual selection, be it inter-sexual or intra-sexual, is not the only context in which moral behaviour appears to play a signalling role. A sexual partnership is only one among many kinds of cooperative endeavours important over the course of a human lifetime. As Barclay observes, signalling via competitive altruism may be expected whenever:

...individuals need to cooperate with others, yet cannot or will not form partnerships with all other group members. The most altruistic individuals may attract the best (or the most) cooperative partners (2004: 217).

The many receivers problem boils down to this: there are many possibilities regarding to whom moral behaviour as a signal may be directed – potential mates, sexual rivals, social competitors, coalition partners – and the available empirical work supports several of these possibilities, not just the first. The many receiver problem does not imply that SSM is wrong, only that
it is likely to be just one part of a more general signalling-based account of moral behaviour.\textsuperscript{163}

The many receivers problem for SSM might be partially ameliorated by noting that sexual partnerships are not always formed by direct interaction between one would-be partner and the other. For example, Apostolou's (2007) review of marriage practices in almost 200 modern hunter-gatherer societies highlights the influence of parents and other close kin on individuals' mating decisions. Apostolou's work suggests that some signalling behaviours by individuals seeking mates may directed at the parents of potential mates rather than at the potential mates themselves. Congruent with this suggestion is Bliege-Bird and colleagues' observation that older hunt leaders among the Meriam emphasise:

...the value of demonstrating one's character to potential in-laws (elder males and females) in order to impress the parents of the "best" (hardest-working) girl to accept (or choose) you as a husband for their daughter" (Smith et al. [2003: 253]).

In addition, it should be recognised that the many receivers problem is not all bad news for SSM. Realising that moral behaviour as a signal may be directed at receivers other than potential mates can dissolve some apparent difficulties for the view. I suggested earlier in this chapter (see §5.2.1) that it was a mistake to suppose SSM predicts low levels of moral display in pre- and post-reproductive individuals. Seeing SSM as part of a larger signalling story helps make clear why this is so. Pre- and post-reproductive individuals may not need to worry about attracting mates, but there are plenty of social problems such individuals do have to confront and which signalling via moral behaviour may well help to address, such alliance formation and maintenance. This point seems particularly valuable in squaring SSM with the observation

\textsuperscript{163} Ronald Noe (2006: 4-5) points out the relative abundance in the biological literature (theoretical and experimental) of "partner control" models of the evolution of cooperation, which focus on how best to interact with a given partner. The paradigm case of such a model is the Iterated Prisoner's Dilemma. Noe contrasts these models with "partner choice" models, which, as the name suggests, focus on choosing partners with whom to interact. Partner choice is undoubtably important in real-world interactions but models incorporating it are under-represented in the biological literature, and experiments testing such models are also rare. Developing SSM would be one way to help redress the imbalance Noe has highlighted.
that individuals in settled parenting partnerships nevertheless engage in costly moral displays. Part of the explanation is surely that such displays help the individual retain their current mate (choosing to stay together is a kind of mate choice), and another part of the explanation probably has something to do with attracting other sexual partners (either to 'trade up,' so to speak, or simply to engage in a sneaky extra-pair copulation). However, I suspect that the largest part of the story about why such individuals continue to conspicuously display their moral qualities centres on the importance of non-sexual relationships in human social life and the efficacy of moral displays in gaining one entry into profitable partnerships and alliances.

One may well wonder at this point what in particular sexual selection can contribute to a more general signalling-based account of morality. Should we, like Mary Jane West-Eberhard, subsume sexual selection under the more general heading of “social selection,” meaning “differential reproductive success (ultimately, differential gene replication) due to differential success in social competition, whatever the resource at stake” (1979: 158). So what, one might wonder, if the resource at stake is sexual partners rather than social allies? Randolph Nesse is another who suggests that a sexual selection-based account of the evolution of morality be assimilated to a more general, social selection-based account:

Mate choices create potent selection forces, but so do choices of relationship partners. The fitness benefits from choosing social partners are more distant from direct reproduction, but they can influence fitness nearly every day and at all ages. If partnerships yield a net gain for both parties, then fitness increases with the increase in the number of others who want you as a partner, at least for the first few partners. If partners vary in value, then fitness will be increased by behaving in ways that increase the number of others who want you as a partner (2007: 148).

I certainly agree that there is a real question to be asked about whether, in the sexual domain, people are doing any more than using general, social-world adaptations for choosing and being chosen as partners. I think, though, that sexual selection does have something distinctive to contribute to a more general, social selection-based account of morality.
To see the importance of sexual selection in particular, it helps to distinguish (as does Miller [2007: 101]) between good genes and good partner signals. Good-making features of partners across a range of contexts (sexual and social) may include generosity, fairness, trustworthiness, bravery, and loyalty. It is, however, hard to see why one would care about the genetic quality of a hunting partner, say, or political ally. In such cases, phenotypic quality seems like all that should matter. It is clear, though, why genetic quality would matter in a sexual partner. I therefore think that sexual selection specifically is needed in order to understand costly, other-benefiting behaviours that signal good genes.

To sum up, the evidence that moral behaviour plays a signalling role indicates this role is played in contexts beyond mate choice, making SSM seem part of a more general, signalling-based account of morality. Conspicuous displays of generosity, bravery, and the like may serve to attract mates, gain entry into profitable coalitions, and deter social and sexual rivals. Within this general account, sexual selection does have a distinctive role to play, namely, that of explaining ‘good genes’ signals. However, SSM (and the more general signalling-based account of which it should be seen as a part) remains hostage to empirical evidence that is, as I have repeatedly stressed throughout this chapter, currently partial and inconclusive.

5.5. Conclusion

This chapter has presented and evaluated the empirical case for SSM. I began by clarifying the empirical commitments of SSM and, in so doing, forestalled two misguided objections regarding, firstly, the development and expression of moral capacities across life-history stages and, secondly, the existence and extent of sexual dimorphism in morality. SSM properly understood does not predict that moral competence will appear only at or shortly before sexual maturity or that moral displays will decline as individuals age beyond their peak reproductive years. Regarding sexual dimorphism in morality, both the predictions of SSM and the empirical findings most likely to be relevant are currently vague. It can, however, be said with certainty that SSM does not
predict stark sexual dimorphism in human morality. Whether subtle dimorphism exists, and if so then what form it takes, is a currently open empirical question.

Numerous studies across several disciplines – in particular psychology, anthropology, and experimental economics – combine to support the claim that generous, fair, and honest behaviours play a signalling role. This is good news for SSM. However, I have argued, these findings suggest that such signalling behaviours are likely directed toward many receivers in addition to potential mates, including potential social allies as well as sexual and social rivals. SSM should therefore be seen as one element in a more general signalling-based account of morality.

This finding is congruent with the those of my previous chapters. SSM is undeniably on to something. The view, however, must be connected up with other work on the evolution of morality, on moral psychology, and on signaling theory in a far more complex fashion than has yet been managed. This chapter, I hope, constitutes a small step in that direction.
Chapter 6. SSM: Metaethical Implications

6.1. Introduction

This chapter continues the project, begun in Chapter 1 (see §1.1.2e), of drawing out the philosophical implications of SSM. I will focus here on the possibility that an evolutionary explanation for our tendency to make moral judgements might inform arguments in metaethics. Specifically, I will concentrate on the attempted Darwinian debunkings of morality by Richard Joyce (2006) and Sharon Street (2006). As mentioned in Chapter 1, both Joyce and Street argue that an evolutionary explanation for our tendency to make moral judgements should lead us to doubt that any of our moral judgements are true. In this chapter, I will present their arguments in detail. In doing so, I have two aims.

My first aim is to tease apart two debunking strategies. Very roughly, one targets the content of moral beliefs while the other targets the causes of moral beliefs. I will elaborate on the difference between the two strategies and on why distinguishing them matters. My second aim in presenting Joyce and Street’s arguments is to set up a discussion of what I call the ‘reliabilist reply’ to Darwinian debunkings of morality. In essence, the reliabilist reply runs as follows: establishing that our tendency to make moral judgements is an adaptation, far from undermining our confidence in those judgements, may

164 Unlike the five preceding chapters, this one will not proceed via Miller’s own work on the central topic, even though Miller has discussed the implications of SSM. In brief, Miller claims that we generate moral judgements in such a way that conflicts among those judgements are likely to be common, so the method of reflective equilibrium is unlikely to succeed, and thus normative ethics is likely to fail (2007: 241). Miller draws on interesting data from current empirical moral psychology but he misunderstands the method of reflective equilibrium. The aim of that method is to achieve equilibrium among our considered moral judgements (Rawls [1971: 47]; see also DePaul & Copp [2006: 599]; Mikhail [2008: 354]). Even if our moral judgements are initially generated in ways that produce many inconsistencies, we may revise or reject some or even many of our initially-generated moral judgements in order to eliminate, or at least minimise, conflicts. The method of reflective equilibrium does not aim to somehow accommodate all of our moral judgements. Thus, even if our mechanisms for generating moral judgements produce a great deal of inconsistency, normative ethics (conceived of as the attempt to achieve reflective equilibrium in the moral domain) need not fail. In sum, while Miller’s work on SSM itself is worthy of much discussion, his remarks on the implications of SSM for moral philosophy do not bear much critical scrutiny, and so will not be considered in detail in this chapter.
instead provide reason to think they are, by and large, true, since it is more fitness-enhancing to hold true beliefs than to hold false ones and selection will therefore favour cognitive mechanisms that are reliable – that track the truth – over those that are unreliable.

Both Joyce (2006: 182) and Street (2006: 126) anticipate essentially this kind of reply to their debunking efforts. Both offer responses to the reliabilist reply but, I will argue, both responses leave something to be desired. I propose to meet the reliabilist reply head-on by identifying the conditions under which it is reasonable to expect evolved cognitive mechanisms to be reliable, then considering how our evolved moral faculty fares in relation to these conditions. The conditions are, in brief, as follows:

1. the environment condition, i.e. the mechanism is operating in an environment relevantly similar to that in which it evolved
2. the information condition, i.e. information is not high cost, thus it is not adaptive to employ a cheap, error-prone mechanism
3. the error condition, i.e. asymmetrical error costs are unlikely to have selected for systematic bias in the mechanism
4. the tracking condition, i.e. the function of the mechanism is to track features of the agent's environment

I will elaborate upon these conditions later in the chapter. For now, I will simply foreshadow the material to come. I argue that our evolved moral faculty plausibly meets the first three conditions above. I take this to partially explain the appeal of the reliabilist reply. However, I go on to argue that if SSM is true, then the final condition fails. If SSM is true, the function of our moral faculty is not to track features of the world but rather to get us to signal mating-relevant features of ourselves to potential sexual partners. In this way, I claim, the truth of SSM would undermine the reliabilist reply to Darwinian debunkings of morality.

---

165 And, taking into account the many receivers problem presented at the end of the previous chapter (see §5.4), also to get us to signal socially-relevant features of ourselves to potential interaction partners – allies and rivals – more generally. The function remains a signaling one, though, rather than a tracking one.
6.2. Evolutionary Debunking Arguments

In this section, I will first identify the target of evolutionary debunking arguments, then go on to consider two such arguments, one from Richard Joyce (2006) and another from Sharon Street (2006).

Evolutionary explanations of morality do not seek to explain the existence of specific philosophical moral theories, such as Mill’s utilitarianism, Kant’s deontology, or Aristotle’s virtue ethics. Rather, the explanatory target is the species-typical human tendency to think in moral terms. Correspondingly, evolutionary debunking arguments are not targeted, in the first instance anyway, at philosophically sophisticated moral theories. Rather, the target is so-called ‘folk morality.’ Evolutionary debunking arguments, as I understand them, seek to establish that there is something fundamentally mistaken in the species-typical human tendency to think in moral terms. Of course, in order to get such an argument going, it is necessary to define the target more precisely by specifying just what it is to think in moral terms. I have already presented Joyce’s view that an important feature of moral judgements, as far as giving an evolutionary genealogy of such judgements goes, is their “practical clout” (see §2.2.3b). We see moral judgements as inescapably authoritative, making them effective motivators of action. I will now consider another aspect of the psychology of moral judgement-making. It is this aspect, I think, that makes moral judgements vulnerable to the debunking arguments I will discuss below.

As was mentioned in Chapter 1 (see §1.2.2e), the view that morality is, in some sense, objective, is common among philosophers. Don Loeb (2007) warns that, even though such philosophers may claim to be speaking for the folk, we should not simply assume that their views are representative of the folk view of morality. For a start, Loeb claims, “it seems illegitimate to attribute much metaethical theory to most people” and, moreover, evidence gathered by observing folk moral talk and practice “may well take us in more than one direction” (2007: 472, 473). Loeb’s points are well-taken. To gain an
understanding of folk morality (i.e. the target of the debunking arguments I will discuss below) we need to consider empirical work.

Disappointingly little of the abundant empirical work on moral judgement deals directly with “the psychology of metaethics,” as Geoffrey Goodwin & John Darley ([2008]; [forthcoming]) put it. To remedy this, Goodwin & Darley have investigated whether people not only hold various first-order moral beliefs but also second-order moral beliefs – *metaethical* beliefs – about the objectivity or subjectivity of their first-order moral beliefs. Following Geoffrey Sayre-McCord ([1986]; [2009]), Goodwin & Darley say that to see morality as objective is to:

...hold that the truth conditions of moral claims are mind-independent in the sense that a moral claim can be true without reference to the subjective states of the individual making the judgement, and without reference to the conventions of any group of people who are making the moral judgement.\(^{166}\)

Goodwin & Darley’s subjects were not baldly asked whether they considered themselves to be objectivists. Rather, they were questioned about cases of moral disagreement in a way designed to reveal their metaethical stance (2008: 1343-1351). Subjects were given a collection of statements on various topics, including ethics, geography, and music.\(^{167}\) Subjects were first asked to rate their level of agreement or disagreement with each statement. For each statement, subjects were then asked if they thought there was a correct answer about whether that statement was true (a yes/no question). Finally, for each statement, subjects were asked whether, if another person disagreed with them over the truth of that statement, then either that person or the subject him/herself must be wrong.

---

\(^{166}\) It is worth pointing out that this way of understanding what objectivity amounts to meshes well with the way that Shaun Nichols spells out the notion as it appears in the work of various philosophers:

To claim that an action is *objectively* immoral is to claim that the action is wrong “as it is in itself” and not in relation to subjects. Cashing out “as it is in itself” is no pleasant undertaking, but we might focus on the negative claim... If morality is objective, then being morally wrong is *not* determined by a relation subjects bear to the action. Thus, to claim that an action is objectively immoral is to claim that the action is wrong “as it is in itself” and not in relation to other subjects (2004: 5).

\(^{167}\) One statement used in the study was ‘Robbing a bank in order to pay for an expensive holiday is a morally bad action.’ Another was ‘Massachusetts is further north than California.’ A third was ‘Classical music is better than rock music.’
Notice that one could sensibly answer ‘yes’ to the first question above but ‘no’ to the second only if one thought that the truth of the relevant statement was somehow mind-dependent. For example, in the case of statements about music, one might answer in this way if one thought that which musical styles are better than which others depends on an individual’s tastes. If one thought that the truth of the relevant statement was not mind-dependent, as most people presumably do in the case of statements about geography, then one should answer ‘yes’ to both questions. Answering ‘yes’ to both questions for a given statement was thus taken as evidence of being an objectivist about that statement.168

Loeb would no doubt be pleased by Goodwin & Darley’s results. The investigators themselves sum things up nicely: “[L]ay metaethical systems are not monolithic” (2008: 1354). Most subjects gave objectivist responses to most ethical statements. However, not all subjects treated ethical statements as objective, nor were all ethical statements treated as equally objective. Subjects who self-reported a belief in God were more likely to treat ethical statements as objective than were atheistic subjects. Ethical statements about the wrongness of moral transgressions were more likely to be treated as objective than were statements about the goodness of morally exemplary acts (see Goodwin & Darley [forthcoming] for more detail on this point).

One hopes Goodwin & Darley’s findings are only the beginning of a fruitful research program aimed at understanding the psychology of metaethics (or ‘folk metaethics,’ as it might be called). For now, though, those findings can be taken as provisional support for the following claim. Often when making moral judgements, ordinary moralisers believe themselves to be describing the way the world is, independent of the beliefs, preferences, or interests of those making the judgments. Morality, like geography but unlike music, is typically seen as a domain in which truth is objective in the sense defined above.

168 As Goodwin & Darley (2008: 1358) put it, “objectivism is measured by combining participants’ judgments about the status of a particular belief that they hold [i.e. possibility of a correct answer] with their judgments about the mistakenness of someone who disagrees with that belief.”
My discussion of evolutionary debunking arguments below will be conducted against the background of this objectivist view of morality. I am aware that some responses to these arguments "go constructivist" about morality (Levy [2006: 582]; see also Carruthers & James [2008]). Such responses will not be discussed here. For one thing, I see adopting a non-objectivist metaethic in response to attempted evolutionary debunkings of morality as burning the village in order to save it. More pragmatically, space does not permit me to adequately present and critique constructivist-style responses. Seeing off the reliabilist reply will be enough to occupy me in this chapter.

Having identified the target of evolutionary debunking arguments, I turn now to consider in detail two attempted Darwinian debunkings of morality. The first is Joyce's (2006) argument, the second is from Street (2006).

Joyce begins his debunking argument with the "belief pill" thought experiment (2006: 179). He invites us to imagine discovering that we have beliefs about Napoleon only because we have been slipped a pill that causes us to form beliefs about Napoleon. Without the pill, we would not have any beliefs about Napoleon at all. This discovery amounts to finding out that our 'Napoleon beliefs' (so to speak) are the products of a process in which the truth of those beliefs does not figure. Napoleon's being short, Corsican, defeated at Waterloo – indeed, his having existed at all – have nothing to do with our believing these things about him. Joyce claims that, once we discover this fact, it would be epistemically irresponsible of us to continue believing the things we do about Napoleon. We should, rather, cease to believe those things (which is not to say 'disbelieve'), pending further evidence. In short, discovering that our beliefs about Napoleon are products of a process in which the truth of those beliefs does not figure renders our beliefs about Napoleon unjustified.

169 Of course, there are other reasons one might adopt some kind of non-objectivist metaethical view. I am concerned here with just one small region of the vast argument-space that is metaethics.
Joyce claims that our moral beliefs are analogous to the Napoleon beliefs in the thought experiment above, insofar as our moral beliefs are the products of a process in which their truth does not figure. Natural selection has, in effect, slipped us a moral belief pill: “[w]ere it not for a certain social ancestry affecting our biology, we wouldn’t have concepts like obligation, virtue, property, desert, and fairness at all” (Joyce [2007: 181]).

Joyce defends the view that the human tendency to deploy these and other moral concepts – to judge in moral terms – is an adaptation that evolved in response to the problem of sustaining fitness-enhancing but fragile reciprocal relationships ([2006: 140-41]; see also §3.1.1a). I will not reproduce Joyce’s empirical case here, having already touched on much of the material he covers. I will focus on the metaethical upshot he takes his evolutionary explanation of morality to have.

It is worth pointing out that, were it not for a certain ancestry, we wouldn’t have all kinds of concepts: had we not been shaped by evolution so as to perceive the range of colours that we do, we would not have the concepts mauve and puce, for instance. Joyce’s key claim is not just that our moral faculty is the product of evolution, but that the truth or otherwise of moral beliefs did not matter in the evolutionary processes that produced that faculty. He writes:

Whether we assume that the concepts right and wrong succeed in denoting properties in the world, or whether we think that they suffer from a referential failure that puts them on a par with the concepts witch and ghost, the plausibility of the hypothesis concerning how moral judgement evolved remains unaffected (2007: 183).

Joyce’s point is that the evolutionary explanation for our tendency to make moral judgements need only mention nonmoral properties: the property of being fitness enhancing and the property of being motivationally efficacious are two pertinent examples. Joyce recognises a response that might be made, namely, that moral properties are identical with, or supervene upon, certain of the nonmoral properties that do get mentioned in the evolutionary genealogy of morality. It must be asked, he admits, “whether the genealogy of moral
judgement, though couched in non-moral terms, nevertheless implicitly "involves" the existence of moral facts" (2006: 187).

It is at this juncture that Joyce invokes Gilbert Harman’s (1977) argument. It is worth taking a moment to lay out Harman’s argument and to consider just how it meshes with Joyce’s debunking efforts.

Harman invites us to imagine two scenarios. In one, a physicist sees a vapour trail in a cloud chamber and makes the judgement ‘there goes a proton.’ In the other, a woman sees some children burning a cat for fun and makes the judgement ‘burning cats for fun is morally wrong.’ Harman says that the best explanation for the physicist’s judgement will involve, at some point, the existence of protons (1977: 6). He questions whether the best explanation for the woman’s judgement will involve, at any point, the existence of moral wrongness. When it comes to explaining her moral judgement, or indeed moral judgements in general, Harman notes that:

…it would seem that all we need assume is that you have certain more or less well articulated moral principles that are reflected in the judgements you make… There does not ever seem to be any point to explaining someone’s moral observations by appeal to what is actually right or wrong, just or unjust, good or bad. It always seems to be more accurate to explain moral observations by citing facts about moral views, moral sensibility. (1977: 7, 22)

Harman is here raising the possibility that the best explanation for the making of moral judgements may appeal to nothing “about the world over and above” the psychology and socialisation of moral judgement-makers (1977: 7). If that were actually the case, Harman claims, moral theories positing the existence of properties like wrongness, injustice, fairness (etc.) would be in trouble. This is because Harman accepts what Geoffrey Sayre-McCord identifies as the “Explanatory Criterion” for justified belief: “[a] hypothesis should not be believed if the hypothesis plays no role in the best explanation we have of our making the observations that we do” (1986: 267).

The challenge Harman imposes on the believer in moral properties is to show how such properties do in fact play a role in our best explanation for
why we make the moral judgements that we do. Harman’s challenge thus targets the *content* of moral beliefs, that is, beliefs about the existence and the instantiation in particular cases of certain properties: rightness, wrongness, obligation, and the like. Harman in effect demands to be shown how these properties earn their way into our ontology by helping to explain our moral judgements. If this demand cannot be met, then we will not be justified (according to the Explanatory Criterion) in admitting moral properties into our ontology. As a result, we would not be justified in holding particular beliefs about the instantiation, in this or that case, of some moral property, nor would we be justified in holding more general moral beliefs and theories that are committed to the existence of moral properties.\(^{170}\)

It is worth asking at this stage just how Harman’s challenge meshes with Joyce’s debunking efforts. I confess to some uncertainty on this point. One way to see things would be as Joyce simply adopting Harman’s argument. On this way of seeing things, Joyce’s debunking effort consists in adding two things to Harman’s argument. First, he builds on the psycho-social explanation for moral judgement Harman merely sketches, extending it back in time beyond the lifespan of any individual moraliser and into the deep history of our species, then supports it with evidence from various scientific disciplines. Second, he combines Harman’s argument with an argument against moral naturalism: any attempt to reduce moral properties to natural properties will, Joyce argues, fail to “accommodate the sense of inescapable practical authority with which moral claims appear to be imbued” (2006: 190).\(^{171}\)

Another way to see things would be as Joyce attempting to maintain his initial belief pill analogy and using Harman’s challenge to illustrate why he must consider proposed reductions of moral to nonmoral, natural properties. In Harman’s case, such proposals must be considered because, if one succeeds,

---

\(^{170}\) It seems worth pointing out that, if moral properties play no role in the best explanation for our moral judgements, it is unclear how else we could get evidence for their existence. An adherent of a Divine Command Theory of morality might claim to have evidence for the existence of moral properties via the word of God but, revelation aside, there seems no way in which we might gain evidence for the existence of moral properties other than by way of our moral judgements.

\(^{171}\) Harman, by contrast, thinks his own challenge can be met (1977: 132).
then moral properties satisfy the Explanatory Criterion. In Joyce’s case, such proposals must be considered because, if one succeeds, then moral truth does feature (sneakily, so to speak) in the evolutionary genealogy of moral judgement and hence the belief pill analogy breaks down. Seen in this way, there is a clear parallel between Harman’s and Joyce’s work but the two are not offering one and the same argument. Where Harman’s argument proceeds via discussion of what properties we should admit into our ontology (targeting the content of moral beliefs), Joyce’s argument depends on a claim about what processes of belief-formation undermine beliefs so formed (and hence targets the causes rather than the contents of moral beliefs). Both Harman’s and Joyce’s arguments have epistemological conclusions (concerning the justification or otherwise of our moral beliefs) but, seen in the way I have just suggested, they reach those conclusions by different routes.

To help make clear the difference between the two kinds of debunking strategies I am trying to tease apart, I will present Sharon Street’s (2006) Darwinian debunking argument, which focuses on the causes of moral beliefs in order to undermine those beliefs.

Street’s target is “evaluative realism,” the view that “there are at least some evaluative facts or truths that hold independently of all our evaluative attitudes” (2006: 110). Street includes moral judgements under the general heading of “evaluative judgements.” For example, the moral realist according to Street would hold that the judgement ‘Hitler was evil’ is true “independently of any stance that we (or Hitler) might take toward that truth” (2006: 111). Moral truths are thus, for Street’s moral realist, like the truths of physics and chemistry: they hold regardless of whether we believe them or approve of them or are even aware of them at all.

Street does not claim that moral realism thus construed is the folk view of morality. In fact, her argument is explicitly aimed at contemporary philosophers who hold realist moral views and claim those views to be compatible with our current best science (2006: 109). However, if the
empirical studies presented above on objectivism are to be believed, then Street’s debunking argument will also work against folk morality.

Street’s argument begins with a claim about the causal history of our moral judgements:

[T]he forces of natural selection have had a tremendous influence on the content of human evaluative judgements... Different evaluative tendencies... can have extremely different effects on a creature’s chances of survival and reproduction... There [has] been overwhelming pressure in the direction of making those evaluative judgements which tended to promote reproductive success (2006: 113-114).

I will elide the empirical details in Street’s argument. As was the case when presenting Joyce’s argument above, much of the relevant material has already been discussed in previous chapters. I will again focus on the supposed metaethical upshot of our tendency to make moral judgements being an adaptation.

Street claims that discovering our moral judgement-making tendencies to have been so heavily influenced by our evolutionary history confronts the moral realist in the following way:

The challenge for realist theories of value is to explain the relation between these evolutionary influences on our evaluative attitudes, on the one hand, and the independent evaluative truths that realism posits, on the other (2006: 109).

If there were no relation between the evolutionary influences on our moral judgements and the independent moral truths, then barring “an incredible coincidence,” Street says, “most of our evaluative judgements are likely to be false” because “the historical push of natural selection on the content of our evaluative judgements has nothing to do with evaluative truth” (2006: 121, 125). On the plausible assumption that we are not justified in believing things we know have only the remotest chance of being true, our moral beliefs would thus be unjustified.
If, on the other hand, there were some relation between the evolutionary influences on our moral judgements and the independent moral truths, the question would be what kind of relation holds between the two. Street thinks the only option for the moral realist here is to claim that:

...the relation is this: in ways roughly analogous to the ways in which we were selected to be able to track, with our non-evaluative judgements, facts about such things as fires, predators, and cliffs, so we were also selected to be able to track, with our evaluative judgements, evaluative facts (2006: 136).

Street calls this the “tracking relation” (2006: 125). If this relation holds between the evolutionary processes that shaped our moral tendencies and the realist’s independent moral truths, then those processes:

...should not be viewed as distorting or illegitimate at all. For the evaluative judgements that it proved most selectively advantageous to make are, in general, precisely those evaluative judgements which are true (2006: 126).

I wish to pause at this point to note the similarities between Joyce’s belief pill analogy and the first horn (the ‘no relation’ horn, so to speak) of the dilemma Street presents the realist. In both cases, the crucial claim leading to the conclusion that our moral beliefs are unjustified is that the processes that have led to our holding such beliefs are not processes that track moral truth. In fact, both Joyce’s belief pill analogy and the no relation horn of Street’s Darwinian dilemma can be cast in what Guy Kahane (forthcoming) calls the “general form” of a debunking argument. Kahane quite sensibly writes:

All beliefs have a causal explanation. But if someone decided whether or not to believe that p by flipping a coin, her belief would surely be unjustified; there is simply no connection whatsoever between this means of forming a belief and the truth. What matters here is not whether a belief was shaped by a process that is literally random but whether it was shaped by a process that tracks the truth.

Street notes that the realist could claim that evolutionary influences have pushed us away from the independent moral truths – an ‘anti-tracking’ relation, one might say – but of course this would make it even less likely than in the no relation case that any of our moral judgements match up with the realist’s independent moral truths.
Kahane refers to processes of belief formation that do not track the truth of the beliefs so formed as “off-track” processes. Debunking arguments, Kahane claims, have the following form:

Causal premise: S’s belief that \( p \) is explained by \( X \)
Epistemic premise: \( X \) is an off-track process
Therefore, \( S \)’s belief that \( p \) is unjustified

In Joyce and Street’s case, the relevant subjects in the causal premise are ‘all humans,’ the relevant beliefs are ‘all moral beliefs,’ and \( X \) stands in for the suitably-detailed and empirically supported evolutionary genealogy of the human moral faculty. In Joyce and Street’s case, the epistemic premise claims that evolution is not a truth-tracking process with respect to moral truth.

Notice that if the epistemic premise is true, it wouldn’t matter to this debunking argument if moral properties were in fact identical with some naturalistically respectable properties. There could be, for all we know, reduction relations of the appropriate kind between moral and natural properties, but as long as the case can be made that the process by which our moral beliefs are formed is not a reliable one, the debunking argument laid out above works. In this way, the debunking strategy pursued by Street (and embodied in Kahane’s general debunking argument) side-steps the broadly metaphysical issues raised by Harman’s challenge.

To reject the evolutionary debunking argument as laid out above, one would have to either deny the causal claim (an option I will not consider here), deny the claim that beliefs produced via off-track processes are unjustified (an unpromising option), or deny that evolution/natural selection is an off-track process with respect to moral truth. I will focus on this last option, which is what I call the ‘reliabilist reply.’ Both Joyce and Street anticipate essentially this kind of reply to their debunking arguments. Joyce recognises the appeal of the thought that “perhaps the process of natural selection is likely to yield true

---

173 I am not suggesting that moral naturalism plus scepticism is an attractive (or even a coherent) package of views. The point is rather that the metaphysical and the epistemological issues are, to an extent, independent of each other.
beliefs” (2006: 182), while Street identifies the ‘tracking’ account as her realist opponent’s best bet. I turn now to consider the reliabilist reply in detail.

6.3. The Reliabilist Reply

In this section, I will try to make clear why the reliabilist reply seems like an attractive response to evolutionary debunking arguments. I will then discuss Joyce and Street’s responses to the reliabilist reply, noting some problems with those responses. Next, I will do the reliabilist reply what looks like a favour by carefully laying out the conditions under which it is reasonable to expect evolved belief formation mechanisms to be reliable. By making the reliabilist reply more precise, though, I hope to make it easier to reject. That will be my task in the next and final section of this chapter.

It is, apparently, tempting to suppose that holding true beliefs is more likely to be fitness-enhancing than holding false beliefs. Several quotes will illustrate the prevalence and prominence of this kind of view:

Natural selection guarantees that most of an organism’s beliefs will be true, most of its strategies rational (Dennett [1987: 75]).

Creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die out before reproducing their kind (Quine [1969: 126]).

Darwinian selection guarantees that organisms either know the elements of logic or become posthumous (Fodor [1981: 121]).

In the orthodox naturalistic view, correspondence [truth] is a resource, or a general purpose fuel for success in dealing with the world (Godfrey-Smith [1996: 172]).

The adaptive value of believing truths over falsehoods and acting accordingly should be obvious. It is no great task to imagine how a hominid hunter-gatherer would be far more likely to survive and reproduce if he guided his behaviour by true rather than false beliefs about what substances are nutritious, which organisms are dangerous, the kinds of things potential mates

174 The first three quotes are cited in Stephens (2001: 161).
find attractive, and how others would be likely to react to various actions he himself might perform.

The thought that true beliefs are more likely to be fitness-enhancing than false beliefs leads naturally to the thought that evolved belief formation mechanisms will be likely to generate more true beliefs than false ones. That is, it might be supposed that evolved belief formation mechanisms are likely to be reliable.

It is important to realise that the sense of ‘reliability’ in play here is the epistemologist’s sense rather than the practical sense. The relevant sense of reliability is spelled out by Alvin Goldman:

[A] process, method, system, or what have you... is reliable if and only if (1) it is a sort of thing that tends to produce beliefs, and (2) the proportion of true beliefs among the beliefs it produces meets some threshold, or criterion, value (1986: 26).

It is also important to realise that the issue so far as the reliabilist reply goes is not whether evolved belief formation mechanisms in general are reliable (as Dennett can be read as claiming in the quote above), but whether our evolved moral belief formation mechanism in particular is reliable. As Brandon Fitelson & Elliott Sober say:

[O]ur cognitive mechanisms are reliable on some subjects, unreliable on others, and of unknown reliability on still others. We should divide our beliefs into categories and associate a characteristic degree of reliability with each of them (1998: 116-17).

For example, the psychological machinery that allows us to attribute mental states to others seems like a good candidate for being a reliable

---

175 William Rottschaefer claims in his discussion of evolution and morality that “moral justifications [can] make appeal to identifiable reliable mechanisms for attaining moral ends” but Rottschaefer thinks ‘reliability’ is a matter of “how well a mechanism works in doing what it is designed to do” (1998: 204, 208). Our faculty for moral judgement may work reliably in this sense, fulfilling some adaptive function(s), but so does our spleen: this kind of reliability (and the corresponding kind of justification) has nothing to do with truth.

176 Alvin Plantinga ([1993]; [2002]) mistakenly lumps all our cognitive faculties in together when arguing that evolutionary theory and philosophical naturalism together imply that human cognitive faculties are unreliable.
cognitive mechanism (we are quite accomplished mindreaders), while the psychological mechanism that allow us to pick faces out of background detail seems less so (we are prone to over-detect faces in the patterns we perceive), and our faculty for probabilistic reasoning seems like a good candidate for being unreliable (we are subject to several kinds of systematic bias in such reasoning). The reliabilist reply should thus be understood as the claim that the evolved elements of our psychological make-up comprising our moral faculty operate in such a way as to produce more true moral beliefs than false ones.\textsuperscript{177}

I will now turn to Street and Joyce's responses to the reliabilist reply. While my sympathies ultimately lie with the debunkers, I nevertheless think there are certain weaknesses in their responses.

Street says that it is mysterious on the tracking account how making true moral judgements helped our ancestors:

\begin{quote}
According to the tracking account, making certain evaluative judgements rather than others promoted reproductive success because these judgements were true. But now let's look at this. How exactly is this supposed to work? Exactly why would it promote an organism's reproductive success to grasp the independent evaluative truths posited by the realist? The realist owes us an answer here. It is not enough to say, "Because they are true." We need to know more about why it is advantageous to apprehend such truths before we have been given an adequate explanation (2006: 129-130).
\end{quote}

Street points out that merely grasping truths on its own confers no advantages in terms of survival or reproduction. Indeed, she notes, it may be disadvantageous in those terms to grasp certain truths, given the cost of doing so and their irrelevance to the business of surviving and reproducing.\textsuperscript{178}

The tracking account is inferior, Street claims, to what she calls the "adaptive link" account, according to which:

\begin{quote}
...tendencies to make certain kinds of evaluative judgements rather than others contributed to our ancestors' reproductive success not because they constituted perceptions of independent evaluative truths, but rather because
\end{quote}

\textsuperscript{177} I will leave aside here the question of just where the threshold for reliability lies.

\textsuperscript{178} Street's example is truths about "the presence or absence of electromagnetic wavelengths of the lowest frequencies" (2006: 130).
they forged adaptive links between our ancestors’ circumstances and their responses to those circumstances, getting them to act, feel, and believe in ways that turned out to be reproductively advantageous (2006: 127).

It is not at all mysterious on the adaptive link account how making moral judgements helped our ancestors: it did so by making them more likely to perform certain kinds of fitness-enhancing behaviours (as discussed in §2.2.3b and §3.1.1a).

On the tracking account as Street presents it, the benefit of making moral judgements somehow stems from grasping or apprehending moral truths, while on Street’s favoured adaptive link account, the benefit comes about because moral judgements prompt adaptive behaviours. An advocate of the tracking account is likely to cry foul at this point, and rightly so. The natural response to Street on behalf of the tracking account, I think, is that it is not merely grasping truths but grasping truths and guiding one’s behaviour by them that is supposed to be selectively advantageous. The thought motivating the tracking account (as laid out earlier in this section) is that if an organism is going to guide its behaviour by judgements at all (rather than just having a range of hardwired responses to various stimuli), then it will be better off if guided by true judgements than by false ones. Street stacks the deck against the moral realist by allowing a link between belief and behaviour in her favoured account but omitting such a link in her statement of the tracking account.

Reading Street charitably, she may be best understood as demanding to know how, on the tracking account, the truth of moral beliefs figures in the explanation for why the behaviours those beliefs prompted were fitness-enhancing. After all, Street admits that it is clear why an organism would benefit from grasping truths about fires or lions: “the fire might burn it to a crisp; the predator might eat it up” (2006: 130). Street surely does not think an organism that believes, correctly, it is being charged by a flaming lion is for that reason alone better off. The true believer only benefits if it uses its belief to guide its behaviour in a certain way (in this case, getting out of the way).
That this might be the heart of the issue is suggested by a footnote Street makes to her comparison between the tracking account and the adaptive link account. She writes:

In order to explain why it proved advantageous to form judgements about the presence of fires, predators, and cliffs, one will need to posit in one's best explanation that there were indeed fires, predators, and cliffs, which it proved quite useful to be aware of, given that one could be burned by them, eaten by them, or could plummet over them (2006: 160-161).

While it is clear how the truth of true beliefs about dangers figures in an explanation of why it would be beneficial to hold and be guided by such beliefs, it is far less clear that the truth of beliefs about morality must figure in an explanation of why it would be beneficial to hold and be guided by those beliefs. So long as moral beliefs prompt various adaptive behaviours, one can wonder, does it matter whether they are true or false? Should it not matter, then it is unclear what supposing the beliefs to be true adds to the explanation for why holding and being guided by them was adaptive. If Street’s objection to the tracking account is not really that it is mysterious how merely grasping truths could be adaptive, but rather that it is unclear what role (if any) moral truth plays in the explanation for why morally-motivated behaviours were adaptive, then her response to the reliabilist reply is very much like that offered by Joyce.

Joyce explicitly considers externalist process reliabilism as a view about justification on which it may be tempting to see the evolutionary genealogy of moral beliefs as a vindicating one. That view holds that “a belief is justified if and only if it is the product of a process that reliably links beliefs with truth” (2006: 212).³ Joyce claims that, by this view’s own lights, if we had:

...an empirically confirmed hypothesis of how [a] belief-formation mechanism works which does not require that the beliefs be even approximately true, we would have to conclude that any such...beliefs are products of an unreliable process (2006: 215).

³ Exactly what it means for a process to “reliably link beliefs with truth” is not further explicated, which will become important in what follows.
The evolutionary genealogy of morality is just such an empirically confirmed hypothesis about our moral belief-formation mechanism, Joyce points out: “nowhere does the evolutionary hypothesis [about morality] assume that moral beliefs are or were true” (2006: 215). The similarity between Street (as construed above) and Joyce here is hopefully apparent: both stress the apparent superfluity of moral truth in explaining why we make moral judgements. What, then, can be said about this response to the reliabilist reply? As I mentioned above, I am on the side of the debunkers in this debate, but I do think there is one thing to be said – part clarification, part criticism – of the response Joyce (and Street, as construed above) offers.

Joyce’s argument that our moral beliefs are produced by an unreliable process is criticised by Walter Sinnott-Armstrong (2006). The basis of Sinnott-Armstrong’s criticism is a distinction between actual reliability and counterfactual reliability:

A belief process is reliable in the actual world when it yields few, if any, false beliefs in the actual world. A belief process is reliable in counterfactual worlds when it would yield few, if any, false beliefs even in very different possible worlds (2006: 44).

Joyce’s observations regarding the evolutionary genealogy of our moral faculty may well establish, Sinnott-Armstrong allows, that our moral faculty is counterfactually unreliable. All else being equal, that is, we would have evolved to make moral judgements even in worlds where there exists nothing corresponding to our moral concepts. Our moral faculty would be unreliable in those worlds. Sinnott-Armstrong points out that the counterfactual unreliability of our moral faculty does not imply the actual unreliability of that faculty (2006: 44). We may (luckily) inhabit a world that contains moral properties and the process by which we generate moral judgements may (luckily) produce more true moral beliefs than false ones. Sinnott-Armstrong then offers the following analogy:

---

180 Sinnott-Armstrong is actually criticising Joyce’s (2001) evolutionary debunking argument, but the key claims from that earlier work carry over into the later (2006) work I have been discussing so far.
If thermometers work in our world, we seem justified in trusting them here, even if they fail in other possible worlds. Similarly, if we live in a world with moral facts, and our moral beliefs are actually reliable in this world, then the fact that our moral beliefs would be false in a very different possible world without moral facts need not make our moral beliefs unjustified in this world (2006: 44).

To show that our moral faculty is actually unreliable, Sinnott-Armstrong says, Joyce must “assum[e] that there are no moral facts in the actual world” (2006: 44). Joyce is not entitled to this assumption, since whether our world contains moral properties is one of the things at issue in his disagreement with would-be vindicators of morality. Sinnott-Armstrong concludes that Joyce has not given us reason to accept the claim that our moral faculty, here, in the actual world, is unreliable.

To evaluate Sinnott-Armstrong’s criticism of Joyce, I think we need to distinguish between two epistemological views: Goldman’s (1979) process reliabilism and Nozick’s (1981) reliable-predictor theory. The former is a view about justification. It says (roughly) that S’s belief that \( p \) is justified if that belief is the product of a process that produces more true than false beliefs. The important feature of the belief-formation process here is reliability, as defined by Goldman in the quote given earlier. Reliable-predictor theory, by contrast, is a view about knowledge. It says that S’s true belief that \( p \) does not count as knowledge if S would have believed \( p \) even if \( p \) were not the case. The crucial feature of the belief-formation process here is not reliability in Goldman’s sense but what has been called sensitivity to the truth (see Goldman [2009]).

Having drawn the distinction between Goldman’s and Nozick’s senses of reliability, it is easier to see what is going on in Sinnott-Armstrong’s criticism of Joyce. The evolutionary genealogy of morality shows that our moral faculty is insensitive to moral truth – unreliable in Nozick’s sense, or ‘counterfactually unreliable’ as Sinnott-Armstrong puts it – but that does not show our moral faculty is unreliable in Goldman’s sense. As briefly mentioned above, Joyce doesn’t elaborate on the nature of reliability when he talks about processes reliably linking beliefs with truth. Perhaps he (and evolutionary
debunkers more generally) could simply be Nozickians about reliability and rest content with the conclusion that our moral faculty is insensitive to moral truth. This seems debunking enough, since it preserves the belief pill analogy. Even though Nozick was interested in defeaters to a belief’s counting as knowledge, unreliability in his sense also very plausibly serves as a defeater to justification.

Showing that our moral faculty is insensitive to truth leaves us uncertain about whether that faculty is actually reliable or not. To put the reliabilist reply’s lights out, we need more. We need to show that it is not reasonable to expect our moral faculty to be actually reliable. I think it is possible to do this. The way to do it, I think, is by laying out the conditions that would have to be met in order for us to reasonably expect our evolved moral faculty to be actually reliable and showing that not all of those conditions are met.

Laying out the conditions under which it is reasonable to expect evolved cognitive mechanisms to be reliable will be done below. Considering how our evolved moral faculty fares in relation to those conditions will take up the final section of this chapter.

6.3.1. Conditions for Reliability

Earlier in this chapter (see §6.1), I laid out four conditions that I claimed had to be met by an evolved belief-formation mechanism in order for us to reasonably expect that mechanism to be reliable (in Goldman’s sense). The four were the environment condition, the information condition, the error condition, and the tracking condition. In this section, I will elaborate on these conditions. A useful way to view the conditions is as potential defeaters for the claim that a particular evolved cognitive mechanism is likely to be reliable. That is, if one or more of the following conditions fail to be met, we should not expect reliability.
6.3.1a. Environment

Cognitive mechanisms that are reliable under certain circumstances can be unreliable under others. Research on so-called ‘fast and frugal’ heuristics provides a good illustration of this point.

Peter Todd & Gerd Gigerenzer ([1999]; [2000]) describe how psychological mechanisms for making various kinds of judgements can be fast – requiring little computation – and frugal – requiring little information – but nevertheless accurate, performing as well as more complex and informationally-demanding methods. The key to fast and frugal mechanisms’ success, Todd & Gigerenzer say, is “ecological rationality,” by which they mean taking advantage of the way information is structured in the environment (2000: 736). The specific details need not detain us.181 The important point for current purposes is Todd & Gigerenzer’s claim that:

...evolution would seize upon informative environmental dependencies [i.e. the way information is structured in the environment] and exploit them with specific heuristics if they would give a decision-making organism an adaptive edge (2000: 736).

It is worth stressing that the “adaptive edge” here is supposed to come from the accuracy of the judgements generated by the heuristics. We have not slipped into using the everyday sense of ‘reliable.’ Todd & Gigerenzer’s somewhat surprising finding is that fast and frugal but ecologically-rational mechanisms of belief formation can produce true beliefs (or at least, a high proportion of true beliefs to false ones). The idea is put quite nicely by Peter Carruthers:

The idea is that, in connection with any given heuristic, there will be a range of different environments and environment types within which that heuristic will operate with a significant degree of reliability. And we can think of the heuristic as having been selected (by evolution, by individual learning, or through the success of a particular culture) to operate in those environments (2007: 187-188).

181 For an example of how a very simple heuristic (the “recognition” heuristic) along with environmental information structuring can deliver quite astonishingly reliable judgements, see Goldstein & Gigerenzer (2002).
While an ecologically-rational belief-formation mechanism's being fast and frugal doesn't force it to be unreliable, there is a trade-off that such mechanisms must make: "what works to make quick and accurate inferences in one domain may not work well in another" (Todd & Gigerenzer [2000: 736]). Fast and frugal mechanisms sacrifice generality. Put differently, such mechanisms are vulnerable to error when operating in conditions different to those under which they were selected.

Amos Tversky & Daniel Kahneman ([1974]; see also Kahneman et al. [1982]) have shown how heuristics that are reliable under one set of conditions can be unreliable under others. These researchers do not put their findings into evolutionary context or consider different environments in which a given heuristic may be deployed, but the general lesson they teach is applicable here.

For example, the "availability" heuristic "assess[es] the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind" (Kahneman et al. [1982: 11]). This heuristic is likely to be reliable when highly probable events or members of large classes are brought to mind more easily than improbable events or members of small classes. This may often be the case but, as Kahneman et al. note, it is not always so. For instance, when events that are unlikely can nevertheless be easily and vividly imagined – as in the case of disasters that may befall adventurous expeditions – the chance of such events actually happening will be overestimated by the availability heuristic (1982: 13). This heuristic is (and by extension, heuristics generally are) reliable within a certain context but unreliable outside of that context.

The environment condition, then, states that in order for it to be reasonable for us to expect an evolved cognitive mechanism to be reliable, the mechanism must be operating in the same (or a relevantly similar) environment to that in which it was selected. Environmental mismatch should leave us unwilling to suppose that the mechanism is, in its new setting, reliable. This condition is particularly applicable when the cognitive mechanism in question is a quick-and-dirty heuristic, but even when that is not
so, the fact that a cognitive mechanism is being deployed in a novel environment should give us pause, at least until we have reason to think the mechanism will 'cope' (so to speak) with the novelty.

The idea that information is structured in environments has been helpful in presenting this condition. Information itself is of central importance in the next condition to be discussed.

6.3.1b. Information

The benefits of a reliable belief-formation mechanism must be weighed against the cost of building such a mechanism. Depending on the resultant balance, it may be better to have a more error-prone (i.e. less reliable) but less costly mechanism. A justly famous piece of research on hunting behaviour in frogs can be used to illustrate this point.

Lettvin et al. (1959/1968: 251-252) described the way in which several neural structures in a frog's eyes and brain contribute to successful prey capture. Frogs have four different structures that, it turns out, individually register contrast (i.e. the presence of sharp boundaries in the frog's visual field), convexity (i.e. whether such boundaries are curved), movement (i.e. whether boundaries are moving within the visual field), and dimming (i.e. changes in light intensity across the visual field). In concert, these structures lead a frog to shoot out its tongue at anything within its visual field that registers as a small, dark, moving dot. Lettvin et al. called this arrangement a "bug perceiver" (1959/1968: 258).182

While I do not mean to suggest that frogs ever form the belief that a small, dark, moving dot is a bug, this example is relevant in the following way. We can suppose that the function of the mechanism described above (the predatory strike as well as the neural structures) is to capture bugs and we can

182 Further work has since shown that things are even more complex (see Prete [1999: 141-142] and references therein). The complexities, while fascinating, do not matter for current purposes.
suppose that frogs using such a mechanism capture many bugs, but we must admit that many things other than bugs will also be struck at: anything that registers as a small, dark, moving spot will be targeted. By tightening up its targeting parameters, a frog could increase the proportion of strikes that might nab meals to strikes that merely waste time and energy. This may sound good, but the costs of tightening up those parameters must be weighed against the gain from doing so. If wasted strikes are negligibly expensive in terms of time and energy but it is very expensive to rewire the four neural structures in such a way that strikes are triggered only by bugs, then rewiring is just not worthwhile. To put the things less starkly, there will be some point beyond which the cost of increasing the bug to non-bug strike ratio outweighs the benefit of doing so. Beyond that point, there will be selection against frogs that invest in more accurate but on balance more costly “bug perceivers.”

The general lesson of the example above is that not only error but also accuracy has a cost. James Sage, talking specifically about evolved human cognitive faculties, provides a partial accounting of the costs of reliability:

Truth-reliable cognitive faculties come at a high price: (i) the brain requires oxygen, calories, and cooling, (ii) calculating detailed inferences (even with minimal data) requires considerable time and concentration, (iii) accessing information from past experience requires extensive storage capacity and retrieval pathways, (iv) identifying relevant information requires multi-level sorting subroutines, (v) ranking desires and goals requires extensive deliberation and reflection, and (vi) utilising “detectors” (and other perceptual inputs) requires precision and acuity (2004: 101).

It is worth pointing out that the cost of obtaining the information needed by a reliable belief-formation mechanism can vary. Some information will be relatively readily available while other kinds will be yielded only stingily by an organism’s environment. To use Kim Sterelny’s (2003) metaphor, informationally “transparent” environments may allow for reliability on the cheap, while in “translucent” or “opaque” environments, the costs of gathering information may make reliability prohibitively expensive. In the latter kinds of informational environments, evolved belief-formation mechanisms can be adaptively unreliable: making frequent errors is less costly than gathering and processing the information needed to avoid them.
The information condition, then, states that in order for it to be reasonable for us to expect an evolved cognitive mechanism to be reliable, the cost of accuracy must not be so high relative to the cost of error that it would be adaptive to settle for an error-prone but cheap mechanism.

This condition has examined the cost of accuracy relative to the cost of error. The next condition to be discussed focuses in on errors.

6.3.1c. Error

Adaptive unreliability can result when asymmetrical error costs generate selection for systematic bias in a belief-formation mechanism. To appreciate this point, it helps to start with a distinction between two kinds of error:

There are two very different ways in which an inferential system may get the wrong answer. One way is to infer that $p$ is the case when $p$ is not the case. Following standard practice, I will call these mistakes false positives. The other way is to infer that $p$ is not the case when in fact $p$ is the case. These are the false negatives (Stich [1990: 61]).

Having drawn this distinction, Stich points out that there are cases in which one kind of error is potentially far more costly in terms of fitness than the other. For example, compare eating a poisonous plant in the mistaken belief that it is safe to not eating a safe plant in the mistaken belief that it is poisonous. Or, compare hiding in the mistaken belief that there is a predator nearby to not hiding in the mistaken belief that there is no predator nearby. In each case, a false positive (poisonous when not poisonous; predator when no predator) results in a small cost (a few calories forgone; a little time and energy wasted) but a false negative may be disastrous.

If error costs are asymmetrical, Stich observes, selection will favour belief-formation mechanisms that over-generate the less costly type of error so as to reduce the chance of a potentially disastrous error of the other type. Such mechanisms may be unreliable, but, Stich says:
...natural selection does not care about truth. It only cares about reproductive success. And from the point of view of reproductive success, it is often better to be safe (and wrong) than sorry (1990: 62).

There is an important complexity neglected in Stich’s better-safe-than-sorry story. Stich considers beliefs in relation to single behaviours but beliefs will often be relevant to many different behaviours. Even in Stich’s simple example, we might suppose that believing a plant to be poisonous leads the believer (imagine her to be a distant ancestor of ours) to not only refrain from eating it, but also from collecting it to trade for a stone tool or to pay off a debt or (perhaps) to share around the campfire as a way of displaying her generosity and gathering skill.

Showing that error costs are asymmetrical is fairly easy when the mechanism in question is tightly tied to a single behaviour. To take a very rudimentary example, bivalve molluscs have a “shadow response”: the dimming of ambient light, as would be caused by a predator above, causes them to close their shells (Warrant & Nilsson [2006: 189]). Of course, not all or even most shadows are cast by predators, so the shadow response generates many false positives. These mistakes lose the mollusc a mere moment of filter feeding, though, which is cheap in comparison to the potential cost of a false negative, which may lose it its life.

When the outputs of a mechanism are relevant to many different behaviours, it is harder to determine whether or not error costs are asymmetrical. This is the case when a mechanism outputs beliefs, which are what Sterelny (2003: 30-31) calls “decoupled representations,” meaning “internal cognitive states which (a) function to track features of the environment, and (b) are not tightly coupled functionally to specific types of response.” It is worth stressing (as does Sterelny [2003: 34]) that the coupled/decoupled distinction is a matter of degree. Representations fall along

---

183 I am not suggesting that molluscs form beliefs about the presence of predators in response to sensing a shadow. The errors in the example are inappropriate behaviors, not false beliefs. The example still serves to illustrate, I hope, the point about asymmetrical error costs.
a continuum, with some being relevant to few behaviours (at the limit, just one) and others being relevant to many.

If a cognitive mechanism outputs highly decoupled representations, errors will repeatedly lead the agent astray across a range of situations. In order for such a mechanism to be adaptively unreliable in the way Stich imagines, the sundry costs of a false positive (for instance) would have to be compensated for by the benefit of avoiding a false negative. This is certainly possible (e.g. if false negatives are invariably fatal). The point is that, in considering whether asymmetrical error costs may have led to the evolution of adaptively unreliable cognitive mechanisms, we must consider where the outputs of a given cognitive mechanism fall along the coupled/decoupled continuum. I think it is safe to suppose that Stich’s better-safe-than-sorry lesson will be particularly applicable to cognitive mechanisms that generate coupled representations.

The error condition, then, states that in order for it to be reasonable for us to expect an evolved cognitive mechanism to be reliable, it must be the case that asymmetrical error costs are unlikely to have generated selection for systematic bias in the mechanism.

The notion of tracking featured briefly in this condition, as part of the definition of a decoupled representation. Tracking itself takes centre stage in the next condition.

6.3.1d. Tracking

The reliabilist reply is essentially the claim that evolved cognitive mechanisms are likely to be reliable precisely because they are adaptations. Notice, though, that the mere fact that selection has been operating on a cognitive mechanism would provide any kind of assurance at all that the mechanism is reliable only if selection has progressively shaped that mechanism to be better and better at tracking the truth (within the mechanism’s domain of operation). I will try to illustrate this idea by way of several examples.
It seems plausible to suppose that we have an innate folk physics that allows us to (among other things) estimate the trajectories of flying objects in our environment. It also seems plausible to suppose that selection over many generations has led to us getting ever better at matching our estimates to the typical behaviour of such objects. Tracking truth in the relevant domain is what the elements of our psychology subserving folk physics have been designed to do. The same can be said of evolved belief-formation faculties concerned with less concrete features of our world. That is to say, the kind of tracking I am interested in here need not be visual tracking. For example, take our capacity to mindread. Here again, it is plausible to suppose that there has been selection for the ability to get things right when it comes to forming beliefs about others’ mental states. Tracking the truth in that domain is the function of the psychological mechanisms subserving mindreading.

If, however, a belief-formation mechanism has been shaped by selection to perform some function other than tracking the truth (in a given domain), then the fact that the mechanism is an adaptation for something should not itself incline us to think it will be reliable. I think it is useful at this point to consider discussions of the evolutionary origin of religious belief. Several hypotheses are prominent in such discussions: a by-product hypothesis (e.g. Boyer [2002]; Atran [2002]), the view that religion is a meme (e.g. Dawkins [2006]; Dennett [2007]), and the idea that religion is in some way a group-level adaptation (e.g. Sloan Wilson [2002]; [2005]). I will focus on last of these hypotheses, not because I think it is right but because it nicely illustrates the point I am trying to make with the tracking condition.¹⁸⁴

Consider the belief that there is a sacred order to the world and the belief that we will experience supernaturally-orchestrated punishment or reward after death. David Sloan Wilson advocates a group selection-based evolutionary explanation of religion, according to which the function of such beliefs is “to promote cooperation among group members, and to prevent passive freeloading and active exploitation within the group” (2005: 422).

¹⁸⁴ For current purposes, I am agnostic on which if any of these view about religion is right.
Suppose that Sloan Wilson is right. In that case, the fact that our religious beliefs are the outputs of an evolved cognitive mechanism should in no way increase our confidence that such beliefs are true. This is because tracking the truth about (for instance) the existence of sacred order or an afterlife is no part of the job that (ex hypothesi) selection has designed the religious belief-formation mechanism to perform.

With some examples on the table, I will now try to say something more general about truth-tracking as I understand it. I take there to be a difference between merely being responsive to some feature of the environment, on the one hand, and, on the other, tracking the truth. There is a perfectly legitimate sense in which we can say mosquitoes track CO2 gradients in the air (see §1.2.2c). Truth-tracking as I understand it, however, is a more cognitively sophisticated affair. It is a matter of generating representations of the way the world is that correspond with the actual state of the world. In presenting the tracking condition, my claim is as follows. Discovering that a cognitive mechanism that generates such representations is an adaptation should incline us to think that mechanism is reliable only if selection has acted upon it so as to progressively refine the fit between the representations it generates and the way the world is (or at least, tends to be).

Unlike the previous conditions, failing to meet the tracking condition is not by itself a defeater for the claim that it is reasonable to expect the cognitive mechanism in question to be reliable. It may be that, even though the function of the mechanism is not to track truths of a certain kind, the mechanism nevertheless manages to do so. Whether this proviso applies must be assessed on a case-by-case basis, and I will consider below whether it applies in the case of evolved moral cognition.

The tracking condition, then, states that in order for it to be reasonable for us to expect an evolved cognitive mechanism to be reliable, either the function of the mechanism must be to track the relevant truths or there must be reason to think the mechanism tracks such truths despite that not being what it was designed by selection to do.
This concludes my presentation of the conditions that must be met in order for it to be reasonable to expect reliability of an evolved belief-formation mechanism. In the next section, I will consider how our evolved capacity to form moral beliefs fares in relation to these conditions.

6.4. Reliability and Moral Cognition

In this section, I will do some work for the reliabilist by showing how the first three conditions laid out above are plausibly met in the case of evolved moral cognition. My aim in doing so is not to defend the reliabilist reply but rather to understand its appeal. I think the fact that evolved moral cognition plausibly meets these conditions may partly explain its appeal. If we want to knock the supports out from under the reliabilist reply, my discussion shows that we are best advised to throw our weight against the tracking condition.

The environment condition stated that in order for it to be reasonable for us to expect an evolved cognitive mechanism to be reliable, the mechanism must be operating in the same (or a relevantly similar) environment to that in which it was selected. Comments like Leda Cosmides & John Tooby’s (1997) “our moderns skulls house a Stone Age mind” and Kenan Malik’s (1998) even more evocative gloss on the Evolutionary Psychologists’ claim – “we are Stone Age men living in a space age world” – may lead us to suppose that there is a drastic mismatch between the environment in which our moral faculty evolved and that which we currently inhabit. This would be a mistake.

There are certainly differences between our environment and that of our distant ancestors. Technological level and complexity of social structure stand out as two of the most dramatic. Still, many of the adaptive problems posed to our distant ancestors by their environment are posed to us by ours. Focusing on moral cognition in particular, we still face the problem of managing social interactions, including but not limited to reciprocation, cooperation, cheater detection and punishment, and reputation management. In these respects, our modern environment resembles that in which our moral faculty was designed by selection to function.
At this point, one may well object that a principled way of deciding when two environments are relevantly different is missing here. After all, our ancestors in the distant past had to eat to survive – as we do now – and their environment contained edible matter – as does ours now – but we do not have mammoth on our menu and they did not have Big Macs on theirs. Should the two environments be counted as the same or different, as far as the environment condition is concerned?

I suggest that calling two environments ‘relevantly different’ (in this context) be understood to mean ‘different in a way that would lead us to expect evolved solutions to adaptive problems that worked in one environment to fail in the other.’ Sticking with the food example, it is apparently the case that hummingbirds in urban areas will return again and again to drink from bird feeders filled (by well-intentioned nature lovers) with artificial sweetener solution and will, satiated by zero-calorie sweetness, starve to death (Fleming [1996]). In the modern urban environment, sweetness is no longer a guarantee of high caloric content. While this is a boon for the waistline-conscious human with a sweet tooth, it spells disaster for those hummingbirds whose evolved cognitive machinery is fooled by tasty fakes.

Advances in technology and increases in the complexity of social structures have certainly made the problem of managing social interactions harder for us than it was for our ancestors. For one thing, there is now much greater scope for anonymity in social interactions. Even so, the reliabilist may say, these changes are not so great that our evolved cognitive machinery (in particular our moral faculty) is likely to misfire disastrously like that of the hummingbirds just mentioned.

It is an open question whether or not the environment condition is satisfied in the case of evolved moral cognition. I will assume for the sake of argument that a suitably determined reliabilist could satisfactorily fill in the details lacking in the forgoing discussion.
The information condition states that in order for it to be reasonable for us to expect an evolved cognitive mechanism to be reliable, the cost of accuracy must not be so high relative to the cost of error that it would be adaptive to settle for an error-prone but cheap mechanism.

To assess whether this condition is met in the case of evolved moral cognition, it is necessary to specify the information that is relevant to the operation of our moral faculty. Giving a full specification of the relevant information is a task for the reliabilist. Doing so would amount to giving an account of the nature of moral facts, such that our evolved moral faculty is able to reliably track them. While I am not about to provide such an account, I take it that at least some headway can be made by identifying in general terms the kind of information required as input for moral judgement-making. On any reasonable and naturalistically-respectable account of the nature of moral facts, this will be social information, especially regarding the intentions, motives and interests of other agents, and information about the likely consequences of actions, especially regarding harms and benefits to others.

When considering the costs of gathering morally relevant information, it is crucial to pay attention to the role of other agents in structuring the informational environment. In some cases, the actions of others will make that environment more transparent. As Sterelny notes:

The narrative life of a community – the stock of stories, songs, myths, and tales to which children are exposed – is full of information about what actions are to be admired, and which are to be deplored. Young children’s stories include many moral fables: stories of virtue, of right action and motivation rewarded: of vice punished (2007: 13).

Parents or allies are likely to enhance the transparency of an agent’s informational environment, but competitors, enemies and would-be cheats are likely to push that environment toward opacity. On balance, then, it is hard to say with any certainty that the costs of gathering the information relevant to the operation of our moral faculty are low.
The information condition may nevertheless be met in the case of evolved moral cognition, given that the costs of error are quite likely to be high. For example, mistakenly judging a selfish, lying coward to be generous, honest, and brave may lead one to rely on that person (or recommend him or her to others), which will probably go bad (and spectacularly so) sooner or later. Misjudging one's own moral obligations may also carry a very steep price, extracted in punishment and reputation damage.

While the reliabilist definitely has work to do to show that our evolved moral faculty satisfies the information condition, I take the discussion above to have at least made it plausible that the reliabilist could complete that work.

The error condition states that in order for it to be reasonable for us to expect an evolved cognitive mechanism to be reliable, it must be the case that asymmetrical error costs are unlikely to have generated selection for systematic bias in the mechanism. This is, I think, a tricky condition for the reliabilist to finesse.

There is a plausible case to be made that error costs are asymmetrical for at least some kinds of moral judgement. For example, take judgements about obligations to help others. Helping someone in the mistaken belief that doing so is morally required (rather than merely permissible) may incur a cost in time, energy, risk, or resources but could result in a net gain, since engaging in supererogatory helpfulness is a good way to enhance one's reputation (and perhaps even to signal one's generosity, robustness, or wealth). Even if that kind of mistake is on balance costly, failing to help someone in the mistaken belief that helping was not morally required seems likely to be more costly still: punishment and reputation damage are again relevant here. When it comes to moral judgements about one's duties to help others, we might think it is better to help (and be saintly) than to stand by and be sorry.

The reliabilist's best bet, I think, is to show that the outputs of our moral faculty count as decoupled representations. Recall from my presentation of the error condition above (see §6.3.1c) that Stich's better-safe-than-sorry
lesson is harder to apply the less tightly the outputs of an evolved cognitive mechanism are tied to specific behavioural responses. Deciding where moral judgements fall along the coupled/decoupled continuum is not straightforward, though.

On the one hand, the story I (along with Joyce and Ruse) have told about the function of moral judgements adverts to their motivational efficacy: moral judgements push us to do the things we judge obligatory (right, good, etc.) and refrain from the things we judge forbidden (wrong, bad, etc.). This fact slides moral judgements toward the 'coupled' end of the continuum. On the other hand, if agent $A$ makes a moral judgement like 'helping is obligatory in circumstances $C$', that judgement will guide $A$'s behaviour across a range of situations. It will (obviously) be relevant when $A$ is in $C$, but it will also be relevant when $A$ observes some other agent $B$ in $C$. If $B$ fails to help, $A$ may gossip about $B$'s callousness, try to punish $B$ directly, reject offers of alliance from $B$, and advise everyone else to do likewise. Considerations like this slide moral judgements toward to 'decoupled' end of the continuum. Notice, the more circulation a moral belief gets in an agent’s cognitive economy, the less plausible it is that asymmetrical error costs in any particular scenario will generate selection for systematic bias against the more costly error in that scenario.

I respectfully decline to speculate further on behalf of the reliabilist here. In any case, I think the reliabilist reply ultimately fails because the fourth condition – the tracking condition – is not met, as I will show in the next section.

6.4.1. SSM and the Tracking Condition

The tracking condition states that in order for it to be reasonable for us to expect an evolved cognitive mechanism to be reliable, either the function of the mechanism must be to track the relevant truths or there must be reason to think the mechanism tracks such truths despite that not being what it was designed by selection to do. In the previous five chapters, I have presented the
case that our moral faculty functions to signal mating-relevant qualities to potential sexual partners and, probably, to signal socially-relevant qualities more generally to potential allies and rivals. If this signalling-based account of morality is correct, we would have an evolutionary explanation of morality that does not assign a tracking function to the psychological mechanisms that generate moral judgements. This alone would not mean evolved moral cognition fails to meet the tracking condition. It could be that our moral faculty also functions to track moral truth, or that we manage to track moral truth despite not having been designed by selection to do so. In what follows, I will lay out my reasons for thinking neither of these things is the case. First, though, I will clarify what I mean in saying that our moral faculty serves a signalling function.

According to the signalling-based account of morality, elements of our psychological makeup have been shaped by selection so as to generate moral judgements because those judgements get us to act in ways that effectively advertise our qualities as sexual and social partners to those around us. We also make moral judgements as part of evaluating others as potential sexual and social partners. That is, our moral faculty plays an important role when we are receivers of signals, as well as when we are signal senders. As has been mentioned before (see §2.2.3b), it is important to keep distinct two senses of ‘moral judgement’ when making the claim that the function of moral judgements is a signalling one. Moral judgements *qua* private mental events are not themselves signals (being unobservable) but are an integral part of generating signalling behaviour. Moral judgements *qua* public linguistic events are plausibly signals (or at least, so I have argued in §4.3.3). With these clarifications and qualifications foregrounded, I will go on to consider the two possible ways in which the reliabilist reply might be rescued at this point.

Signalling and tracking need not be mutually exclusive activities. Suppose ascertaining *T*-truths is particularly difficult, requiring a great deal of time and energy and posing some risk as well. In that case, being a reliable
maker of \( T \)-judgements might conceivably serve as a costly signal.\(^{185}\) That is to say, we can imagine an organism signalling certain aspects of its quality to others precisely by reliably tracking truth in some domain. Something like this may well be going on when Arabian babblers compete to perform sentinel duty (as described by Zahavi & Zahavi ([1997]; see §1.2.1). Detecting predators takes time away from feeding, exposes the sentinel to danger and (presumably) requires keen senses. We might call sentinels *conspicuously perspicacious*. I am willing to say that in this case judgements about predators serve both to track features of the world and to signal features of the organism making them.

I do not deny that it is possible to signal certain qualities via reliable tracking, but I doubt that this is what is going on in the case of moral judgement. In Chapter 4, I carefully considered the qualities plausibly signalled by various kinds of moral behaviour, those being resource level, physical condition and dispositions toward cooperation (see §4.2.1). I also carefully considered the costs of various kinds of moral behaviour. The account that emerged does not look analogous to the Arabian Babbler sentinel case. According to SSM, making moral judgements is costly, but not because it is difficult or dangerous to make the judgements themselves. Moral judgement makers are not showing off via conspicuous (and costly) perspicacity. Rather, the costs of making moral judgements lie in the behaviours such judgements help to motivate. To reiterate, signalling and tracking are not mutually exclusive. However, the bare possibility that signalling and tracking might co-occur does the reliabilist no good.

Even if our moral faculty does not have a tracking function in addition to the signalling function assigned it by SSM, establishing that evolved moral cognition fails to meet the tracking condition requires showing that we do not have reason to think our moral faculty allows us to track moral truth despite its not having been designed by selection to do so. In what follows, I will consider two ways in which the reliabilist might try to claim that we track

---

\(^{185}\) cf. Dessalles ([1998]; [2006]) on costly signaling and the evolution of language.
moral truths despite that not being the function of our moral faculty. I will argue that neither way is convincing.

One could imagine a determined reliabilist trying to find some property or cluster of properties instantiated by all and only the actions that we judge to be morally wrong (call it $W$) and some property (cluster) instantiated by all and only the actions we judge to be morally right (call it $R$). If this attempt were to succeed, might the reliabilist not be warranted in claiming that we reliably track $W$ and $R$ with our moral judgements of wrongness and rightness? And if an action’s being right or wrong is just a matter of its instantiating $W$ or $R$ (respectively), might the reliabilist not be correct to say that our we reliably track moral facts?

There are numerous problems with the reliabilist proposal just laid out. For a start, the existence of widespread moral disagreement poses an immediate and substantial stumbling block to identifying $R$ and $W$.\textsuperscript{186} Even if such properties could be found, there would be the question of whether identifying moral rightness and wrongness with those properties is plausible. This question takes us back toward issues in the metaphysics of morality – issues I have tried hard to skirt – but I think such issues can be avoided yet again. The claim that if there were an $R$ and a $W$ then we would count as reliably tracking those properties is, I think, mistaken. To see why, it helps to consider again Sloan Wilson’s proposed evolutionary explanation of religious belief (see §6.3.1d).

As was the case earlier, I am not here supposing an evolutionary explanation of religious belief in terms of group-level benefits is correct. That said, allow me construct a deliberately simplified thought experiment. Imagine that all actions judged to be damning (meaning ‘such as to guarantee one entry into the unpleasant afterlife’), were discovered to share the property of being

\textsuperscript{186} See Fraser & Hauser (forthcoming) for an empirically-oriented discussion of moral disagreement and metaethics.
such as to erode in-group cohesion.\footnote{Or to have had that property in the ancestral environment in which the tendency to form such beliefs evolved.} Would this be grounds for supposing that the psychological machinery that generates beliefs about damnation is reliable? I think the answer here is clearly `no.` The content of the beliefs generated by a cognitive mechanism matters – crucially – when assessing the reliability of that mechanism. In-group cohesion is not what the religious belief in the thought experiment is about: that belief is about damnation and the afterlife. So, it is not enough for reliability that there be \textit{some} feature held in common by everything to which a class of judgements applies: that feature must match up with the content of those judgements.\footnote{The point I am trying to make here is essentially the same one I made when first presenting the tracking condition (see §6.3.1d), about the difference (as I see it) between tracking the truth in a given domain, which is a matter of forming representations of the world that correspond to the actual state of the world, and merely responding in some counterfactually robust way to certain features of the environment, which is something even mindless slime can do.}

Let's return now to the moral case. If SSM and the more general signalling-based account of morality of which it is a part are correct, then $R$ (the property instantiated by all and only the actions we judge to be morally right) may turn out to be something like `being such as to improve the probability of sexual intercourse and/or acceptance into profitable social alliances.' As Miller quite rightly pointed out in \textit{The Mating Mind}, considerations like these are typically not what we have on our minds when we engage in moral behaviour and make moral judgements. While a lot of philosophical and empirical spadework is needed to figure out what we do have in mind when making moral judgements (spadework like that mentioned in §6.2), it is not so hard to rule things out. On no serious analysis is `being such as to improve the probability of sexual intercourse' going to emerge as the property we mean to attribute to actions when we judge them to be morally obligatory. Thus, I take this first way of making the case that we manage to track moral truth despite not having been designed by selection to do so, to be a failure.

In response to the previous five chapters of this thesis, one might concede that we have the capacity to think in moral terms for the reasons...
given in the signalling-based evolutionary explanation but insist that, however we got that capacity, we can now consciously direct its exercise and examine its workings and thereby discover moral truths. On this kind of view, the evolutionary genealogy of morality is only temporarily debunking, or debunking of only some of our moral judgements. Even if we are not now reliable moral judgement-makers, with some effort we could be. For example, Michael Huemer allows that “[e]volution may have endowed us with biases that affect our moral judgements” but claims that “[s]ociobiology can help us identify these biases and so correct for them, thereby improving moral cognition” (2005: 219).189

One important way in which we can correct for evolution-induced biases in our moral judgements, Huemer says, is by paying attention to coherence relations among such judgements. He offers the following analogy:

Suppose a detective interviews six eyewitnesses to a robbery. All of them claim to have seen the robbers drive away in their getaway car. The detective interviews the witnesses separately, giving them no opportunity to confer with each other. Now suppose that two of the six witnesses agree that the getaway car had license plate number X78 41A, while the other four witnesses report four different license plate numbers. In this case, even though most of the witnesses are wrong, the detective could still conclude that the correct license plate number was probably X78 41A. The reason is, in essence, that it is extremely unlikely that even two witnesses would agree on a specific license plate number, in the absence of collusion, unless that number were correct (2008: 379-380).

In this case, we can make a judgement about what the license plate number was that is very likely to be correct, even though the majority of reports received were inaccurate, because the accurate reports cohere with each other. In the moral case, Huemer says, even if we “suppose that only a third of our ethical intuitions [are] accurate,” we can nevertheless expect those intuitions to cohere with each other and so we can be “prima facie justified” in taking the members of the largest coherent subset of our moral intuitions to be “roughly accurate” (2008: 380).

189 Huemer seems to use the term ‘sociobiology’ quite inclusively, meaning it to cover not just the scientific field so named but also evolutionary biology-inspired theorising about humanity more generally.
Huemer's thought experiment is a poor analogy for the case of evolved moral cognition. In the license plate case, we have good independent reason, aside from the eyewitness reports, to accept that there are such things as license plates and that those things have numbers and letters on them. We also have good reason to think that some eyewitnesses – those who were fortuitously situated and clear-sighted – could have accurately read the getaway car's plate. Against this background, we may well use coherence relations among number plate judgements to form a reasonable best guess about what the getaway car's plate said. Independent reason, aside from our moral judgements themselves, is precisely what we lack in the moral case: our only reason to think there are any moral truths at all is our tendency to judge in moral terms. Nor, in the moral case, do we have reason to think anyone is especially well-placed to make accurate moral judgements: the evolutionary genealogy of moral judgement implies that we are all subject to selection's distorting influence. Against this background, it is unclear why a debunker would grant even that just one in three of our moral judgements are true. Notice, finally, that in the moral case there is collusion of a sort: we all share the same evolutionary history. Coherence in the presence of collusion like this, by Huemer's own lights, does not show that the cohering judgements are likely to be true.

Huemer has one more move to make. He claims:

[N]ot all intuitions are equally open to the accusation of bias. For instance, biological evolution would be expected to produce a bias toward favorable evaluations of things that promote one's own inclusive fitness; intuitions that do not imply favorable evaluations of things that promote one's own inclusive fitness are not candidates for being products of this particular bias (2008: 381).

Huemer is right that evolutionary explanations of morality suppose the action of selection to be non-random, to skew our moral judgements in specific directions. He is, however, wrong to think that “[e]thical intuitions that do not conform to the relevant expectations are not open to the charge of being produced by these biases” (2008: 381). His mistake lies in paying insufficient attention to the distinction between explaining why we make
certain particular moral judgements rather than others, on the one hand, and, on the other, explaining why we have the tendency to judge in moral terms at all rather than lacking that tendency entirely. Even moral judgements concerning matters unconnected to inclusive fitness (for example, the judgment that terraforming Mars would be morally impermissible) count as “being produced by” evolutionary forces, insofar as the tendency to make such judgements would not exist were it not for the action of those forces. The causal chain linking any moral judgement made by a modern human to the selective forces operative upon ancestral hominids is going to be long, complicated, and take idiosyncratic turns through personal and cultural history. What Huemer has not done is provide reason to think that chain connects up at any point with moral truth. It thus remains unclear to me why we should think there is anything to correct towards, to home in upon, to more reliably apprehend, when we make moral judgements.

Allow me now to sum up this last section. SSM as an evolutionary explanation of morality assigns a signaling function to the evolved elements of our psychology that allow us to make moral judgements. Playing a signaling role does not exclude also playing a tracking role. I have argued that our moral faculty does not play both roles. Lacking the function of tracking moral truth does not preclude doing so anyway. I have argued against two ways in which I imagine the case might be made that we reliably track (or could reliably track) moral truth despite not having been designed by selection to do so. While I am conscious of the many things that remain undiscussed, I will conclude at this point that evolved moral cognition fails to meet the tracking condition. On that basis, I claim, it is not reasonable for us to expect the psychological machinery that generates moral judgements to be reliable. The reliabilist reply to attempted Darwinian debunkings of morality can now, I think, after receiving an extended hearing, be put away.

6.5. Conclusion

This chapter has considered the philosophical implications of the truth of SSM. I have identified two strategies employed in attempted Darwinian
debunkings of morality. Metaphysical debunking arguments target the content of moral judgments, aiming to show that the properties referred to by such judgments ought not be admitted into our ontology. Epistemological debunking arguments, by contrast, target the causes of moral judgments, aiming to show that such judgments are the products of an unreliable process.

I have focussed on the epistemological debunking strategy and, in particular, on a certain kind of reply to that strategy: the reliabilist reply. According to this reply, the fact (if it be one) that our moral faculty is the product of a long evolutionary heritage should lead us to think that faculty is reliable, that it tracks the truth. In order for it to be reasonable to expect an evolved cognitive mechanism to be reliable, however, certain conditions must be met. These are the environment condition, the information condition, the error condition, and the tracking condition.

The appeal of the reliabilist reply to Darwinian debunkings of morality that employ the epistemological strategy is partly due, I have suggested, to the fact that the first three conditions listed above are plausibly met in the case of evolved moral cognition. Even so, the truth of SSM undermines the reliabilist reply by making it implausible that the tracking condition is met. SSM can in this way contribute to the efforts by Darwinian debunkers.

In this chapter and those prior to it, many empirical questions about the evolutionary genealogy of moral judgment have been left open and many important metaethical issues have been bracketed and sidelined. I am well aware that my discussion is far from conclusive on any of the points I have raised. Even so, my hope is that I have contributed to an ongoing, cross-disciplinary and (when viewed from a sufficiently elevated vantage point) cooperative effort to explain how it is that we came to be moral creatures and what it is that this explanation implies about morality.
Works Cited


Barclay, P. 2004: Trustworthiness and competitive altruism can also solve the "tragedy of the commons". *Evolution and Human Behavior*, 25, 209-220.


Fraser, B. & Hauser, M. forthcoming: The argument from disagreement and the role of cross-cultural empirical data. *Mind and Language*.


Gurven, M. 2004: To give or not to give. Behavioral and Brain Sciences, 27, 543-583.


