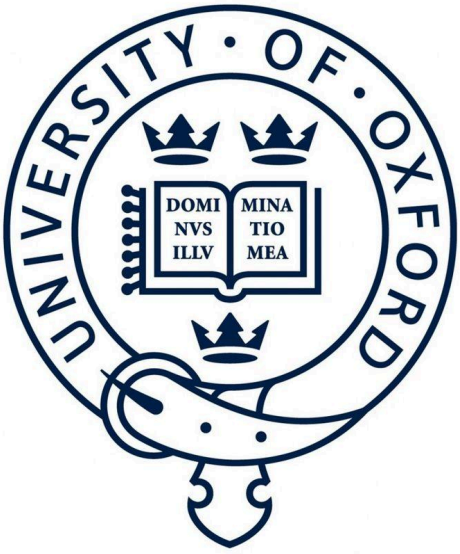
Bio-Constructivism

Towards an Organisational Account of Virtue



Milan Ney

St. Hilda's College

A thesis submitted for the degree of

Doctor of Philosophy in *Philosophy*

Trinity 2025

In memory of Anke Jungclaus

Rest in Power

Acknowledgements

I would like to thank my supervisors, Edward Harcourt and Alison Hills, for their invaluable comments, criticism and support; Bernd Spannuth for his generosity at a crucial time; Nick Pappas, Jesse Prinz and Jonathan Gilmore at the City University of New York for their institutional backing; as well as M.M. and J.M. alongside their children --- J. and J. --- and N.R.-M., of Harlem and N.J., for welcoming me like kin.

I have special gratitude for my parents, Beate Ney and Ernst Jungclaus, who taught me never to know my station, and for Megumi Masanobovna Kaizu, who's been putting up with it all for over a year now.

Table of Contents

ONE: Introduction....6
TWO: Function-based naturalism....10
1. Introducing function-based naturalism....10
2. Meta-ethical naturalisms and the challenge from science....13
3. The threat of 'biologism' or hypostatisation....21
4. Conclusion....40
THREE: Neo-Aristotelian naturalism and its discontents....41
1. Neo-Aristotelian naturalism....41
2. Vital judgements as the 'manifest image' of life....43
3 Arguments for natural-historical judgements....49
4. Conclusion....63
FOUR: From the organisational account of virtue… 64
1. The organisational account of function....64
2. The functionality of character traits....69
3. The regulated functions of character traits....76
4. Regulated functions and the received virtues....82
5. Conclusion....88
FIVE: … to bio-constructivism....89
1. The phenomena of ethical inquiry....90
2. The apparent dilemma between hypostatisation and trivialisation....93
3 Ethical inquiry according to bio-constructivism....104
4. Converging and judging others....126
5. Concluding remarks: Ethical inquiry and natural history, revisited....128
SIX: The virtues of (distant) others....129
1 Virtue-inquiry concerning particular distant communities, and cultural pathologies....130
2 External arguments for universal virtues....138
3 Virtues of moderation....141
4 Virtues of justice....147
5 Common vices and universal virtues....153
6 Conclusion....156
SEVEN: How virtue-inquiry motivates to regulate....158
1. Ethical self-control motivation and the Default Account....158
2. The Ambitious Account....167
3. The normative challenge....185
4. Conclusion....191
EIGHT: Conclusion....192
References....194
Appendix A: Function-Based Naturalisms (A Diagram)....208

ONE: Introduction

Myopia is a defect of eyesight; cruelty a defect of character. Oaks ought to have strong roots and humans ought to be charitable. Human beings can have a *good* or *bad* sense of justice, just as most animals can have a *good* or *bad* sense of smell. There is a striking similarity between the way we describe functions or purposes in living things, and the virtues of human beings. Surely, this insight should allow us to connect ethics and biology! Thus, it could bring us closer to understanding how ethical truth can fit into our world, which in many ways appears to be fully explainable and describable by natural sciences.

There are many views in (meta-)ethics building on this insight. Yet, the most influential among them do *not* link ethics and biology in such a way (e.g. Foot, 2001; M. Thompson, 2008). While they connect ethics to our understanding of purposes in living nature, they divorce that understanding from scientific biology. This is because they want to preserve the autonomy of ethics and avoid (what I will call) *hypostatisation*. As they see it, founding ethics on any notion of function taken from scientific biology would place particular ways of being human beyond ethical criticism. It would declare them to be ethically right without that judgement being supported by the right --- ethical --- kind of reasons. However, through this move, they give up the hope of showing how ethical truths fit into a world explainable by natural science. Instead, they endorse 'liberal naturalism' and deny that the world is exhaustively explainable by natural science. As I will argue, this is a mistake. And anyway, they do not successfully escape hypostatisation either, or so, I will attempt to show.

Given this, I present an alternative account based on the same insight: *bio-constructivism*. I argue that ethics can be conceptually founded on scientific biology while avoiding hypostatisation. At the heart of bio-constructivism is the *organisational account of virtue*. Virtues are character traits which actualise their functions according to the organisational account of function (Mossio et al. 2009, Moreno and Mossio 2015). Function according to this account is a matter of self-maintenance and self-regulation. A trait's function is the contribution it makes to the self-maintenance of the organism as part of which it is maintained, insofar as this function is regulated for.

Bio-constructivism can avoid hypostatisation and maintain the autonomy of ethics because of this emphasis on self-regulation. *Ethical inquiry* itself is a regulatory subsystem. In fact, it is unique as a

general regulatory subsystem. Unlike other regulatory subsystems, it advances self-maintenance on all relevant levels, e.g. on the level of the individual organism as well as of the lineage. Thus, the functions of the character traits regulated by ethical inquiry are in a large part determined by ethical inquiry. Specifically, these functions are determined by how ethical inquiry *would* regulate them if it itself maximally actualised its function. Thus, the best way to determine the functions of these traits is to 'get down to business' and engage in ethical inquiry. Doing so, one makes an effort to have it maximally actualise its function, as one does in all (genuine) ethical inquiry. Ethical inquiry, however, proceeds according to its autonomous logic. Thus, even though correct judgements of ethical inquiry are true descriptions of reality in the same way as any other judgement concerning biological function, ethical inquiry also is involved in the construction of the reality it describes.

This inquiry starts out from judgements based in a particular ethical community and tradition. Thus, according to bio-constructivism, in order to evaluate the actions and character of agents with whom we do not share a tradition, we must engage in a particular form of empathetic ethical inquiry starting from the assumptions of the evaluee's community and trying to uncover mistakes they might have made. We can be reasonably sure that there is some convergence between non-mistaken inquirers on the broad outlines of 'justice', but it is an open question whether there is such convergence on all judgements or on a substantial notion of 'the good life'.

I argue that bio-constructivism is an attractive position, primarily, because it can account for the peculiar logic of ethical inquiry. It can do so without denying that the natural sciences are in principle capable of describing and explaining all that there is, and while maintaining that ethical judgements are true in exactly the same sense as other ordinary and scientific judgements. Additionally, seeing virtues as character traits with organisational functions is epistemically fruitful. It allows us to see analogies between human life and the lives of other animals which can enrich both ethical inquiry and scientific inquiry into the lives of animals. Finally, it supports an interesting if so far speculative account of ethical motivation based in 'autopoetic enactivism', a view on mental content content founded on a notion of function closely related to the organisational account (cf. E. Thompson 2007).

Let me give a chapter-by-chapter summary. In **Chapter Two**, I introduce the notion of *function-based naturalism*. Function-based naturalism encompasses all meta-ethical views that treat virtue or practical

rationality as a species of purpose or function in living nature. I further introduce the dilemma between meeting the challenge from science and avoiding hypostatisation. In order to meet the challenge from science, a naturalistic account must show that the subject matter of ethical inquiry is continuous with the subject matter of (other) natural sciences. In order to avoid hypostatisation, an account must not place any way of being human beyond ethical criticism. I argue that all forms of function-based naturalism found in the literature so far which rely on notions of function based in scientific biology fail to avoid hypostatisation. All of these accounts are based on *evolutionary* notions of function.

In **Chapter Three**, I describe *neo-Aristotelian naturalism*. According to neo-Aristotelian naturalism, judgements about natural purposes and excellences constitute a *sui generis* form of discourse distinct from judgements about 'function' made within biological science. Thus, neo-Aristotelian naturalism can apparently avoid hypostatisation. However, I firstly argue it cannot meet the challenge from science. I consider several arguments to the contrary and reject them. Furthermore, many forms fail to avoid hypostatisation as well.

In **Chapter Four**, I present the *organisational account of virtue* as an alternative form of function-based naturalism which is based on notions of function from scientific biology but avoids hypostatisation. It draws on Moreno and Mossio's organisational account of function. According to the organisational account, a trait T of organism O has the function F iff T does F to some degree, T contributes to the self-maintenance of O by doing F and T's doing F is regulated for by another subsystem of O. I argue that many traits determining human behaviour make contributions to self-maintenance that are incommensurable relative to each other. Which of them are functional depends on which are regulated for. Ethical reasoning itself is a regulatory subsystem. Thus, the organisational account does not preempt ethical reasoning and does not place any features of human life beyond ethical critique.

In **Chapter Five**, I develop the organisational account of virtue into bio-constructivism. I describe the implications of the organisational account of virtue for ethical inquiry. I argue that ethical inquiry cannot be subsumed under scientific inquiry, but follows its own logic. Thus, the challenge of hypostatisation is revolved. Ethical inquiry proceeds through the revision of ethical judgements which the inquirer finds themselves already endorsing. They revise these judgements in the light of a

range of constraints which can be derived from the function of ethical inquiry. These are consistency constraints, thrivingness and shareability constraints.

In **Chapter Six**, I discuss the ways in which bio-constructivism allows us to evaluate distant others and to argue for universal virtues. To do so, I draw on analogies between human beings and other animals. I explore virtues of moderation and the need to regulate different motivational systems that might otherwise lead to a 'run-away' positive feedback loop; virtues of justice and resolutions of (evolutionary) tragedies of the commons.

Finally, in **Chapter Seven**, I present a speculative way in which bio-constructivism can account for the motivational force of ethical judgements. Doing so is necessary to explain how ethical inquiry can regulate character traits, and thus for bio-constructivism to explain the logic of ethical inquiry. I present a Default Account and an Ambitious Account. According to the former, it is a brute fact that humans typically have a desire to be virtuous. According to the latter, the *truth-conditional content* of ethical judgments according to the organisational account of virtue explains their motivational force. Drawing on Evan Thompson, I argue that intelligent behaviour, intentionality and life are closely connected. All life exhibits (proto-)intentionality. Its environment is disclosed to it in a valenced way, i.e. as good or bad. Good environmental features are those which contribute to self-maintenance. Living beings must be *ceteris paribus* disposed to seek them out. Otherwise, their behaviour would not be recognisably intelligent or intentionally directed at the world. Human beings are disclosed to themselves as part of their environment. In the judgement that one ought to be/act in a certain way, being/acting that way is presented as contributing to self-maintenance. Thus, while such a judgement may *sometimes* fail to motivate, it must *typically* motivate. Otherwise, it could not have the truth-conditional content it has.

TWO: Function-based naturalism

In this chapter, I introduce *function-based naturalism*, a family of meta-ethical views containing both *neo-Aristotelian naturalism* and certain forms of 'evolutionary ethics' or *evolutionary naturalism*. I outline function-based naturalism, describe its appeal and how it is related to other forms of meta-ethical naturalism. In this chapter, I describe the dilemma neo-Aristotelianism attempts to solve, i.e. on the one hand avoiding 'biologism,' or the hypostatisation of certain ways of being human derived from biological science while, on the other, showing that ethical judgements are continuous with judgements of biological science such that the former cannot be dismissed without the latter. In the next chapter, I will show how the neo-Aristotelian solution fails.

1. Introducing function-based naturalism

According to what I shall call *function-based naturalism*, the practically rational human agent is the agent whose capacity for practical rationality is not malfunctioning, i.e. whose capacity to treat considerations as reasons for action is actualising its function(s). The function(s) in question belong to the same kind as what can be called *biological*, *natural* or *vital* functions. Examples of these are the function of the heart to pump blood and the function of eyesight in enabling the lioness to recognise her young. Defenders of different forms of function-based naturalism include: Aristotle, Herbert Spencer (1879), Alasdair MacIntyre (1999), Rosalind Hursthouse (1999), Philippa Foot (2001), William Casebeer (2003) and John Hacker-Wright (2021).[1]

In the following, I first outline function-based naturalism. Then, I describe how it is motivated by an attempt to vindicate ethical discourse in the face of apparent challenges emerging from natural

[1] A start for a more complete list would be: Aristotle, Thomas Aquinas, Herbert Spencer (1879), Ragnar Redbeard (1910 [1896]), Ernst Haeckel (1899), Arthur Twining Hadley (1913), Julian Huxley (1943), E.O. Wilson (see Ruse 1984), William A. Rottschaefer and David Martinsen (1990), James Q. Wilson (1993), Richmond Campbell (1996), Larry Arnhart (1998), James S. Chisholm (1999), Alasdair MacIntyre (1999), Rosalind Hursthouse (1999), Philippa Foot (2001), William Casebeer (2003), John Collier and Michael Stingl (1993; Stingl and Collier 2019), Jacob Ross (2019), Parisa Moosavi (2019), John Hacker-Wright (2021) as well as Jennifer Ryan Lockhart and Micah Lott (2024). The list further includes the pioneer of environmentalism Aldo Leopold (2001 [1949]), at least (in different ways) according to the readings of Bryan Norton (2011) and J. Baird Callicot (1982; cf. also Millstein 2024). Such ecological versions of function-based naturalism are plausibly also found in Pyotr Alexeyevich Kropotkin (2022 [1902]) and Murray Bookchin (1982, cf. Dussault 2022). It also includes Theodor W. Adorno according to the readings by Fabian Freyenhagen (2013) and Kieran Setiya (2024). See Appendix A for a diagrammatic overview over different versions of function-based naturalism and related positions in the philosophy of biology.

science. Finally, I describe how it relates to other attempts to meet the same challenge, and thus to other meta-ethical views.

Function-based naturalism gives an answer to the puzzle of how ethical norms and human agents, who are subject to them, fit into the natural world. Inspired by Aristotle, it suggests that ethical normativity is a species of a phenomenon exhibited by all living beings, yet attaching to a peculiarly human faculty. This normativity also encompasses what are ordinarily called 'biological functions', e.g. a heart's function to pump blood. In the following, I will use the term *vital function* for such functions.

Drawing a connection between ethical normativity and vital functions is attractive since both are often described using the same kind of language. A heart fulfilling its function is a *good* heart. A human being who exhibits the virtues has a *good* character. Myopia is a defect of eyesight, cruelty a defect of character and imprudence a defect of reason. A wolf *ought/should* join in with the pack's hunt. A human being *ought/should* perform acts of charity. Furthermore, there are deeper, structural similarities. Both true and vital normativity describe non-actual normativity. That is, according to either it can be that something *ought* to be a way it is not actually. A wolf with a freak mutation preventing her from joining into the hunt still ought to do so. The miser still ought to be charitable.

According to function-based naturalism, ethical normativity is distinguished from the rest of vital normativity through the faculty that it attaches to. Philippa Foot describes this faculty as *practical reason*. This is the capacity to act for reasons. Through practical reason, human beings treat certain considerations as reasons for action, and perform the actions in question. Thus the good human being (in the sense of ethical normativity) is the human being with a good or well-functioning faculty of practical reason (in the sense of vital normativity), i.e. a practically rational human being. *Reasons* are those considerations which a practically rational human being would treat as reasons. An agent *ought* to act a certain way if they would act that way if they were practically rational, i.e. if there is reason for them to act that way.

In principle, function-based naturalists could endorse a conception of practical reason according to which it consists of an isolated 'module' or 'device' in the human mind, e.g. the capacity to engage in explicit practical deliberation. However, given the project of explaining ethical normativity through

continuities between humans and other living beings, function-based naturalists are likely to stress similarities between human beings and other animals. If so, they have to acknowledge that a lot of human behaviour is produced by mechanisms similar to those of our close relatives, the other higher mammals. These animals are not subject to ethical normativity. Hence, they must lack practical reason. If practical consists in a single module, only a small fraction of human behaviour could be evaluated ethically.

It is more plausible to think of the practical reason as consisting mainly of dispositions which are similar to dispositions present in other animals. These may involve, for example, hunger, fear, submissiveness towards conspecifics of higher status, sexual lust, curiosity as well as joy in companionship, play or gaining status. In human beings, these dispositions are transformed by the presence of certain other capacities. This makes them part of practical reason. These capacities may, for example, be the capacity for reflection or ethical deliberation. Suppose I am running away from somewhere out of fear which I feel for a large dog. Other animals, such a cat, may act very similarly and also do so out of fear. However, I am running away *for a reason*. That reason is the dog's being fearsome and dangerous. Plausibly, this cannot be said of the cat. This is since *I* have a capacity which allows me to articulate that I am running away because the dog is fearsome and dangerous. Thus, the practically rational human being will exhibit a range of cognitive, emotional and motivational dispositions. These can be called virtues. Each virtue is a functional trait which partially constitutes practical reason. The practically rational human being is the same as the virtuous human being.

Not all defenders of function-based naturalism explicitly give a name to the faculty which Foot calls 'practical reason'. Hursthouse (1999, passim) calls it "rationality". However, this expression is potentially confusing. It can name a normative standard of evaluation and a property of those who satisfy it as much as a capacity. Some do not use consistent names for the faculty at all. Arnhart (1998: 8), for example, only proposes the following:

> "The good is the desirable, because all animals capable of voluntary movement pursue the satisfaction of their desires as guided by their information about the world. [...] Only human beings, however, can pursue happiness as a deliberate conception of the fullest satisfaction of their desires over a whole life, because only

> they have the cognitive capacities for reason and language that allow them to formulate a plan of life, so that they can judge present actions in the light of past experience and future expectations."

That is, human beings have a set of narrow unique capacities, i.e "the cognitive capacities for reason and language that allow them to formulate a plan of life", which transform the animal capacities within them such that they "can pursue happiness as a deliberate conception of the fullest satisfaction of their desires over a whole life". However, he does not name either the former, narrow capacity nor the latter, transformed faculty. James Q. Wilson (1993, passim) calls one important module which is unique to human beings and transforms other motivational mechanisms into practical reason, "the moral sense". Casebeer (2003, passim) identifies one as the "module for moral judgement".

2. Meta-ethical naturalisms and the challenge from science

In this section, I will discuss the relation of function-based naturalism to other forms of naturalism in meta-ethics. I argue that both are attempts to respond to a challenge to ethical discourse posed by the success of natural science. Furthermore, they all attempt to respond to this challenge by showing that the subject matter of ethical discourse is continuous with that of natural science. However, they do so in fundamentally different ways.

The relation of function-based naturalism to other meta-ethical naturalisms is complicated. Unlike most forms of naturalism, function-based naturalism is not primarily a *semantic* theory. It is not primarily a theory of how ethical judgements get their meaning, e.g. of how they relate to the properties denoted in them. Rather, it is simply a theory about the nature of the subject matter of these judgments. Like many other meta-ethical positions, however, function-based naturalism can be understood as an attempt to vindicate our practice of making ethical judgements and evaluations as well as acting according to them. In an interview in 2003, after the publication of *Natural Goodness*, her major exposition of function-based naturalism, Philippa Foot recounts (Voorhoeve 2003: 33):

> "[I]t was significant that the news of the concentration camps hit us just when I came back to Oxford in 1945. This news was shattering in a fashion that no one now can easily understand. We had thought that something like this could not happen. This is what got me interested in moral philosophy in particular. In a way, I was always more interested in the philosophy of mind, and I still am very interested in it. But in the face of the news of the concentration camps, I thought, 'It just can't be the way Stevenson, Ayer, and Hare say it is, that morality is just the expression of an attitude,' and the subject haunted me."

She further explains:

> "What these theorists tried to do was construe the conditions of use of sentences like 'It is morally wrong to kill innocent people' in terms of a speaker's feelings or attitudes, or of his or her commitment to acting in a certain way. And this meant that, according to these theories, there is a gap between the facts or grounds for a moral judgement and that judgement itself. For whatever reasons might be given for a moral judgement, people might without error refuse to assent to it, not finding in themselves the relevant feelings or attitudes. And this is what I thought was wrong. For, fundamentally, there is no way, if one takes this line, that one could imagine oneself saying to a Nazi, 'But we are right, and you are wrong', with there being any substance to the statement. Faced with the Nazis, who felt they had been justified in doing what they did, there would simply be a stand-off. And I thought, 'Morality just cannot be subjective in the way that different attitudes, like some aesthetic ones, or likes and dislikes, are subjective.' The separation of descriptions from attitudes, or facts from values, that characterized the current moral philosophy had to be bad philosophy."

This quote describes a project of vindicating or redeeming ethical discourse: morality "cannot be" a certain way. Philosophy which did not account for morality being that way "had to be bad philosophy". This makes sense if the task of philosophy is to vindicate a given practice, i.e. to explain it in a way that shows it to be justified. This project of vindicating ethical discourse is one which Foot shares with the likes of "Stevenson, Ayer, and Hare". However, she suggests that they fail in that task. In order to see how, let us first look at the project of vindication itself.

2.1 The project of vindicating ethics

To vindicate ethical discourse means to explain it in a way which shows it to be legitimate. A vindication of ethical discourse may aim at the *content* of ethical discourse, or merely its *form*. To vindicate the content of ethical discourse means showing that particular ethical judgments are true. To vindicate its form means showing that a discourse with the features of ethical discourse is legitimate. The project of content vindication is problematic. There is substantial disagreement regarding which ethical contents are true both between individuals and between cultures. Even judgements which are for us deeply entrenched, such as that sexual violence is always reprehensible, are not universal. The archaic Greeks of Homer's *Iliad* seem only to have condemned sexual violence if the victimised woman was attached to a Greek father or husband. Claiming 'barbarian' women as a prize of war seems to have been seen as a characteristic act of *excellent* human beings. Thus, the project of content vindication must assume that we have a superior insight into the *truth* of ethical discourse. However, it does not provide a reason for this assumption. Rejecting the reprehensible views of the archaic Greeks and supporting this rejection with reasons belongs *within* ethical discourse, not to its external vindication.

In contrast, the project of formal vindication aims to explain certain features of ethical discourse and show that it is legitimate to engage in a form of discourse with these features. There are ways in which ethical discourse resembles paradigmatic attempts to describe and explain the world, both in everyday judgements and in science. We call ethical judgements *true* or *false*. If someone makes an ethical judgement, we can ask them to support that judgement with reasons or otherwise to defend their warrant in making the judgement. We can judge such a defense as successful or unsuccessful. On the other hand, ethical discourse has an intimate link to action. Some ethical judgements produce actions in such a way that no reasoning could intervene between the judgement and action. Suppose I judge that a perfectly rational (or perfectly good) human being would φ in my circumstances. This judgement may or may not produce an intention to φ or the action of me φing. However, if it fails to produce the action, this cannot be due to some additional *reasoning* in which I reached the conclusion not to φ. This requirement does not amount to a form of 'motivational internalism' about normative reasons. For all I am saying here, an agent may completely fail to be *motivated* by a reason, even if they recognise it. However, this failure to be motivated could not be explained as the agent failing to reach the practical conclusion to φ in some reasoning.

Function-based naturalism, in its most plausible forms, is best interpreted as an attempt at formal vindication rather than content vindication.

Formal vindication may have some implications concerning the content of ethical discourse. For example, function-based naturalism appears to support virtue-based accounts of ethics, and thus to exclude the truth of views like utilitarianism. However, an attempt at formal vindication is justified by its success in accounting for formal features of ethical discourse, not in systematising our prior judgements about its content. If an attempt at formal vindication has some patently *un*acceptable content-wise implications, this will count against it. The features making it attractive as a formal vindication must be very strong to overcome our resistances against accepting the content-wise implications.

2.2 The challenge from science

A project of formal vindication may be motivated by a general, skeptically motivated desire to vindicate any discourse at all or some particular challenge. I suggest that Foot along with most other function-based naturalists and modern metaethicists in general is motivated by a particular worry. We can call it the *challenge from science*.

Ethical inquiry resembles scientific inquiry in many ways. It is an effortful attempt to get things right. It relies centrally on arguments. Its conclusions are expressed in declarative sentences which appear to describe the world. However, the academic study of ethics, as practised inter alia by philosophers and theologians, does not look like a *successful* science. It has been as likely to *produce* new disagreements as it is to settle them. Old positions are not disproven, but are regularly re-visited and defended. Thus, resembling scientific inquiry but lacking the hallmarks of scientific success or progress, ethical inquiry seems worryingly similar to failed projects like astrology (the study of the heavenly bodies' effect on human fate), physiognomy (the study of the relation between facial features and character traits) and spiritualist research (the study of the dead's continued existence and their interactions with the living). Some other areas of philosophy besides ethics may be vulnerable to the same challenge, but this is no argument against its force.

The history of science seems to show that the only rational attitude to these projects is to give them up and accept that their purported subject matter does not exist. This is so even though all of these disciplines already appear as developed branches of 'knowledge' in our earliest written records, i.e. in Babylonian clay tablets, even though they have independently emerged in many different cultures (cf. e.g. physiognomy and astrology in ancient China as well as the near universality of ghost beliefs), and even though they all have corresponding non-scientific everyday judgements. There is popular astrology and the common but unconfirmed belief that the full moon causes criminal and other erratic behaviour. There are also common, popular associations between facial features on character traits: the way in which Oscar Wilde has the features in Dorian Gray's portrait reflect the vices he acquires through his life-style make sense to modern readers who are quite innocent of scientific physiognomy. Yet, they are to be rejected as false superstition and pseudo-science. Similarly, so the worry goes, the success of science and the failure of ethics according to the standards set by science show that the attempted study of ethical truth is a *pseudo-science* while everyday ethical discourse rests on *superstition*.

The challenge from science is not just based on a generalisation about the history of inquiries. Human beings belong to the subject matter of natural science. Their behaviour can, in principle, be explained through natural science. This includes the discursive practices in which they engage. With scientific inquiries, a scientific explanation of their development makes reference to their subject matter. For example, atomic physicists make particular judgements in part because of the outcome of certain experiments. The experiments in question turned out the way they did because of the features of reality which these physicists investigate. In contrast, with discourses like physiognomy or astrology, the scientific explanation of the discourse can*not* appeal to the subject matter of the discourse. As the development of the discourse shows, it is not scientific. Thus, if science's claim to explain human behaviour is correct, these discourses can in principle be explained independently of human behaviour.

With scientific discourses, however, the explanation why the discourses in question are rational involves the connection between their subject matter and the discourse. Astrology and physiognomy resemble scientific discourses so much that, if they were justified, seemingly they'd have to be justified the same way. Thus, given that they are non-scientific, they are pseudo-scientific. Ethical discourse also resembles scientific discourse in important ways. Thus, the worry arises that it too, is

unjustified. The challenge from science is analogous to a challenge provided by natural science to folk psychology, which has been described by Paul Churchland (1989).

The challenge that ethical discourse is unscientific and therefore unjustified appears in various guises throughout modernity. For Kant, it appeared as the worry that science may show that human activity, including action and ethical judgement, is only ever motivated by "inclinations" or animal passions and desires, whereas ethical discourse assumes that human beings can act on recognition of the moral law alone. He responds by demonstrating the limits of scientific cognition, and placing the transcendentally free subject outside of them. For the logical positivists like Ayer, e.g., it appeared as the impossibility of translating moral statements into statements about 'sense data', as they thought the statements of successful sciences were. For Mackie (1977), it consists in the "queerness" of purported ethical facts, i.e. in their apparent ability to compel action if grasped, which supposedly makes them unlike any facts countenanced by the natural sciences. For Street (2006), it arises from an evolutionary account of the faculty of moral judgment. Street argues that ethical facts play no role in explaining why we make the judgments we make. Hence, if "realist" meta-ethics were true, we would have no ethical knowledge. For Streumer (2017), it is the essential contestability which fatally sets ethics apart from science. Even if two agents agree on all facts expressible without relying on core ethical concepts, if they have ethical disagreements, it would be inappropriate for them or an observer to dismiss them as 'merely verbal'. In contrast, in science, if two agents disagree on statements expressed using one concept, but agree on all facts that are expressible without that concept, their disagreement is merely verbal.

2.3 Responses to the challenge from science

Most meta-ethical views can thus be seen as attempts to vindicate the form of ethical discourse in the face of the challenge from natural science. The majority of meta-ethical views (but not function-based naturalism) proceed by giving an account of the nature or character of (apparent) ethical *judgements*. Expressivism, the view defended by "Stevenson, Ayer, and Hare" (see e.g. Stevenson 1945, Ayer 1952 [2012], Hare 1952) with whom Foot contrasts her own view, holds that unlike astrology, physiognomy and spiritualist research, the apparent judgements constituting ethics are not at all in the same business as those of successful science. They do not describe the world. Rather, they express "likes or dislikes" or other kinds of preferences of the speaker.

Foot attacks this view for failing to vindicate a central aspect of our practice of making ethical judgements, viz. that we do not consider ethical disagreements as mere "stand-offs". Rather, we act as if at most one party is right. In this important respect, ethics is more like physics or alchemy than it is, say, like the discourses of football fandom or wine appreciation. Thus, Foot intimates in the interview, expressivism fails to vindicate ethical discourse.

Such a thought has been articulated in a more specific way by Enoch (2011: 19) in his *argument from impartiality*. He suggests that it is a platitude that "in an interpersonal conflict, we should step back from our mere preferences, or feelings, or attitudes, or some such, and to the extent the conflict is due to those, an impartial solution is called for." However, this is not how we act, or think we should act, in the case of a conflict between a Nazi and his enemy. Thus, *prima facie* expressivism and similar views cannot vindicate ethical discourse.

Expressivists like Stevenson (1945) have developed sophisticated ways of accounting for these surface features of ethical discourse. However, Foot seems to think that no such sophisticated account can overcome the problem that ethical disagreements are *fundamentally* different from each other. Specifically, according to Enoch, they struggle to provide a convincing, non-arbitrary explanation for why we treat ethical disagreements differently *practically* than disagreements of mere preference.

Expressivist positions can be called *naturalist* in a broad sense. They do not depend on any assertions that are contrary to the absolute claims of natural sciences. However, the term *naturalism* is used more narrowly of another group of meta-ethical views to which function-based naturalism belongs. According to such views, the subject matter of ethics is continuous with the subject matter of science, and ethical discourse is therefore legitimate.

Most forms of meta-ethical naturalism proceed by giving an account of the *semantics* of ethical judgements, i.e. of *how* ethical judgements get their meaning, of their *relation to* their subject matter. According to these accounts, their referents turn out to belong to the purview of the natural sciences even though ethical discourse looks different from natural science. According to *Cornell-style naturalism* (or 'Cornell realism', see e.g. Boyd 1988), judgments refer to parts of the world in virtue of

some complex causal relation between the types of signs featuring in them and some object or natural kind, i.e. some kind in principle capable of being investigated by the natural sciences. This applies to everyday descriptive judgements and scientific judgements as much as ethical judgements. According to *'Canberra plan' naturalism*[2] (or 'moral functionalism', see e.g. Jackson and Pettit 1995), the signs featured in ethical judgements instead refer to their subject matter in virtue of the connection they have with other signs through 'platitudes'. These are judgements that are taken to be self-evident. In this, they are like theoretical terms in the sciences as well as many, though not all everyday concepts. Like these views, function-based naturalism attempts to meet the challenge presented by the success of natural science to ethical discourse by showing that the subject matter of ethics is continuous with the natural sciences. Function-based naturalism specifically aims to show its continuity with biology. The challenger, it argues, cannot consistently deny the legitimacy of ethics without also denying the legitimacy of much of biology, a reputable and successful science.

Thus, function-based naturalism is a view about the nature of those things that form the subject matter of ethics: virtue, the good life and the good human being, right action, etc. Its claims are expressible with first-order ethical concepts like *virtue* and *right action*. Like other meta-ethical views, however, it is largely agnostic on particular ethical controversies. For example, it is compatible with different stances on how an agent should act if presented with the choice to kill one or let many die. Since it can be expressed in first-order ethical concepts, function-based naturalism is compatible with other forms of meta-ethical naturalism, such as 'Canberra plan' and Cornell-style naturalism. However it does not need to rely on any such semantic views to defuse the challenge from science.

This generates one constraint on a successful form of function-based naturalism. It must show that ethical judgements are indeed continuous with judgements in biology. This continuity must be such that if ethical discourse was illegitimate because it is not scientifically respectable, substantial parts of biology would not be either.

[2] The view is called this because it originated from ANU in Canberra (a planned city!) and is at least potentially part of a broader methodological project --- the Canberra plan --- of accounting for any seemingly non-naturalistic concept, such as those describing mental states, in the same way (see Braddon-Mitchell and Nola 2009).

3. The threat of 'biologism' or hypostatisation

Forms of function-based naturalism fall into two categories, *biological naturalism* and *neo-Aristotelian naturalism*. According to biological naturalism, vital normativity can be reduced to other concepts used in the science of biology. All forms of biological naturalism presented in the literature so far have been forms of *evolutionary* naturalism: they reduce function to some property explained in evolutionary theory. (Non-evolutionary biological naturalism is therefore ignored here, for want of anything to discuss.) In contrast, neo-Aristotelian naturalists believe that vital normativity is a *sui generis* mode of discourse. It cannot be reduced to judgements belonging to some other discourse, including (other) parts of the life-sciences. That is, one cannot give an account of the nature of vital normativity which would make it possible to derive judgements of vital normativity, and thus of ethical normativity, from other judgements in scientific biology.

Even though both forms of function-based naturalism have seen many defenders, neo-Aristotelian naturalism has generated substantially more attention. This is likely due what Michael Thompson (2004: 62) calls the "threat of 'biologism'": function-based naturalism may appear to be "a sort of vulgar evolutionary ethics: a system, in any case, which doesn't know how to distinguish a mere 'is' from the genuine moral or normative 'ought'."

I will argue that this objection can be understood as a criticism of *hypostatisation*. Hypostatisation consists in placing ways of being human and human environments beyond ethical criticism even though they are subject to ethical criticism in our discourses. First, however, I will briefly consider the history of such objections.

The objection to ethical views for failing to "distinguish ... 'is' from ... 'ought'" goes back to Hume (1740). However, it has been most influentially applied by G.E. Moore (1922) in criticising Herbert Spencer's (1879) biological naturalism. For Hume, the concern is about the way in which ethical judgements feature in the motivation of action. According to him, no merely descriptive judgement could motivate action in the way that ethical judgements may. Moore's case, however, is based on the 'Open Question Argument'. According to that argument ethical goodness cannot be identical to any natural property, such as promoting evolutionary fitness, since questions like the following will always be open: 'Preferential altruism towards your kin promotes fitness, but is it really good?'.

Moore's argument is highly problematic. It does not appear to rely on any particular features of ethical discourse. On its broadest reading, it would apply to any identity between two properties: 'This phenomenon is a discharge of electricity from a cloud, but is it really lightning?' and 'this liquid consists of H2O molecules, but is it really water?' are, in a sense, open questions. These identities, however, are virtually uncontroversial. On a weaker reading, the argument is only intended to apply to purported *analytic* identities. These are identities which we may know simply by virtue of grasping the concepts used in them. 'This man is unmarried, but is he a bachelor?' may indeed not be an open question in the way 'kin-altruism promotes fitness, but is it really good?' and similar questions are. However, the naturalist need not present the identity of goodness with a(nother) natural property as an identity of this kind.

However, Moore's Open Question Argument is suggestive despite its flaws. It can be seen as related to Kant's (2019 [1785], IV: 448) remark that "we cannot conceive of a [faculty of practical] reason which consciously responds to a bidding from the outside with respect to its judgments". Korsgaard's (1996, passim) "normative question" argument is a successor. According to that argument, a successful account of ethical rightness must explain why we do not treat questions like the following as open: 'sure, acting altruistically towards your kin is right, but still: why act altruistically towards your kin?'. Rather, we take judgements about the right as settling questions about how to act. However, if ethical concepts like *rightness* and *goodness* simply denote properties like promoting fitness, it seems unexplainable why we should treat judgements about the good or the right in this way. It always makes sense to take a step back from them and ask questions like: 'Sure, acting altruistically towards your kin overall promotes fitness better than any alternative, but still: why act that way?' That is, it always makes sense for some form of reasoning or deliberation to intervene between the recognition of the natural fact and action. This reasoning, may, perhaps not take the shape of strict arguments as they are exemplified by mathematics or natural science. However, it involves considering potential reasons, rather than being merely indecisive, weak-willed or torn between different courses of action.

In Street's (2012) reading of Korsgaard, we need a reason to treat some consideration as a reason. Otherwise, what we are doing does not count as treating anything as a reason. The function-based naturalist seems to offer 'it promotes fitness better than any alternative' or some similar

consideration derived from biology as a reason to act altruistically towards kin. However Korsgaard, on Street's reading, insists that for this suggestion to make sense, there must be a reason to treat 'it promotes fitness better than any alternative' as a reason to act. If I have no such reason, at least implicitly, then I do not count as treating 'it promotes fitness better than any alternative' as a reason at all. Street illustrates this with the example of a formerly hyper-conservative Christian. They used to treat the fact that a dance was lewd as a reason not to participate in it. They had an (apparent) reason for doing so, e.g. that lewdness displeases God. After their de-conversion, they may still have a habitual aversion to those displays of sexuality and eroticism they used to call 'lewd'. However, they do not treat lewdness as a reason anymore. This, according to Street, is because they do not treat anything as a reason for treating lewdness as a reason. The function-based naturalist seemingly cannot offer a reason for treating 'it promotes fitness better than any alternative' as a reason. At most, they may insist that treating it as a reason promotes fitness better than any alternative. Even assuming that this is the case, the response would beg the question.

David Enoch (2011: 100ff) makes a similar point with his indispensability and just-too-different arguments. According to the indispensability argument, we are justified in believing that there are ethical facts since doing so is indispensable for practical reasoning, i.e. reasoning that issues in action. If there were no ethical facts for practical reasoning to aim at, there would be no sense in which a practical reasoner could 'get it right'. According to the just-too-different argument, natural facts are "just too different" from the kinds of facts which could fulfill this role. Thus, again, the problem seems to be that the agent can 'take a step back' from facts concerning natural properties in reasoning about what to do. Hence, these facts cannot be what such reasoning aims at.

A related point is also at the core of Bernard Williams' moral thought according to Miranda Fricker (2020). Williams is committed to "ethical freedom". For him, ethical reflection is a matter of finding one's "deepest impulse" (D.H. Lawrence via Williams, see Fricker 2020: 927) in dialogue with others. This is incompatible with deriving ethical truth from any kind of inquiry aiming at universal objectivity. These include universal reason as in the Kantian project, the consequentialist calculus of utility and Aristotle-inspired attempts to derive them from knowledge about the human species. Though Williams and Fricker don't mention this, the 'open question' may be seen as a way to insist on ethical freedom against the "outrageous annexation of the ethical by self-styled universal rationality" (Fricker 2020: 924) and similar universalistic discourses.

Similar criticisms of evolutionary naturalism in particular have also been presented by T.H. Huxley (1894) and Sidgwick (1902), targeting Herbert Spencer, and C.D. Broad (1944), targeted at Julian Huxley (1943). The objection is also related to the Pollyanna objection to neo-Aristotelian naturalism. According to that objection, neo-Aristotelians either must make scientifically unjustified assumption about 'human nature', or they must accept unacceptable conclusions, such as that lying, sexual violence or ruthlessness are virtuous, since they are natural for humans. In other words, according to the Pollyanna problem, neo-Aristotelian/function-based naturalism would require us to accept such traits as virtuous without providing ethical reasons for doing so. The term "Pollyanna Problem" has been coined by Millgram (2009). Versions of it have also been articulated by MacIntyre (2002), Andreou (2006), Woodcock (2006), Millum (2006), Gowans (2008), Odenbaugh (2017) and Chappell (2015, 2024).

This criticism of (function-based) naturalism can be seen as a critique of *hypostatisation*. Function-based naturalism may appear to hypostatise a certain way of being human, i.e. to place it beyond ethical criticism. Seemingly, in any form of ethical discourse that could be made scientifically respectable by function-based naturalism, asking 'sure, this is how a well-functioning human being would act, but still: why act that way?' would be pointless. Hypostatisation of this kind has been criticised by Theodor W. Adorno. Adorno has been recently read as a kind of neo-Aristotelian and thus as a function-based naturalist (see Freyenhagen 2013, Setiya 2024). However, he criticizes function-based naturalists for hypostatising (*vergegenständlichen*) contingent conditions of human life (see Adorno 1966, Whyman 2016, 2017)[3]. He criticises both Aristotle (see Whyman 2017) and the "Social Darwinists" --- including, implicitly, Spencer --- this way.

[3] *Vergegenständlichen*, 'making object-like' is sometimes translated as *reification*. However, it is different from *verdinglichen* 'making thing-like, thing-ify', which Adorno takes from Lukács and which is more commonly translated as 'reification'. *Verdinglichen*, but not *vergegenständlichen*, contains a play on *verdingen* 'to enter into service (e.g. as a servant or labourer)'. *Vergegenständlichen*, but not *verdinglichen* directly evokes the subject/object distinction. *Vergegenständlichen* could be translated as 'object-ify'. However, this would have strongly misleading connotations. Whyman (2017) introduces 'hypostatise' as a translation. Adorno uses "*Hypostatisierung*" in a closely related way. Another related notion in Adorno is *ontology/ontologisation*.. Adorno uses it for views which place ways of being human beyond ethical criticism by treating them as eternal, extemporal truths, rather than by treating them as part of the natural world as investigated by the sciences. Examples for Adorno include Plato and the Catholic phenomenologist Max Scheler.

Vergegenständlichen derives from *Gegenstand* 'object'. In hypostatising ways of being human, (meta-)ethical theories remove them from the sphere of the subject's freedom to some objective realm, which the subject encounters as alien. In fact, Adorno connects such *Vergegenständlichung* to the success of the natural sciences (see Horkheimer and Adorno 1944). The natural sciences, as part of "Enlightenment", are an attempt to understand nature in order to predict and control it. Thus, natural science enhances human freedom. However, with its progress, natural science expands to human beings as part of nature. Natural science gives rise to the capacity of predicting and controlling human behaviour, including ethical deliberation. This capacity is realised on a social scale, e.g. in 'culture industry' and in totalitarian state-apparatuses. Not conforming to the expectations becomes increasingly less thinkable, and increasingly more penalised. Thus, socially mediated "mastery of nature" becomes a "second nature", which is as merciless towards the individual and restrictive of human freedom as original, 'first' nature. The relation of the individual towards this second nature takes on the characteristics of what Adorno calls "myth": It cannot be questioned or freely appropriated; rather it must be submitted to. The conditions under which the individual lives, and the standards she is expected to conform to, are hypostatised.[4]

We can compare this analysis to a dialectic that arises from the success of natural science. Science's apparent ability to explain ethical discourse leads to the challenge from science. Naturalistic accounts presented in response to the challenge from science hypostatise certain standards and conditions of human life, i.e. they imply that they are beyond ethical critique. However, these standards and conditions are in fact subject to critique in our ethical discourse, so these accounts fail to vindicate ethical discourse.. This dialectic differs from Adorno's analysis in so far as for Adorno, different philosophical positions express or reflect different social formations. Thus, a philosophical account that hypostatises certain conditions of human life reflects a social formation that makes ethical critique of them impossible. Still, if any philosophical account of ethical discourse is acceptable for Adorno, it must not affirm such a social formation.

[4] This is, of course, a much simplified discussion of Adorno. Among other things, Adorno appears to fear that human freedom may actually be so diminished and ethical discourse may actually become so deformed as to make criticism of these circumstances impossible. It may become impossible to 'let suffering speak'. Thus, Adorno would disagree that we can treat the possibility of such criticism simply as a desideratum in a meta-ethical theory. In fact, Adorno's complex and difficult literary style may be due to an attempt to recover a nearly lost potential for criticism that resides in language due to its connection to the human life-form (see Freyenhagen 2023).

This worry about hypostatisation can be explained further by contrasting our form of ethical discourse with the one MacIntyre (1981:) sees in "heroic societies". MacIntyre's primary example of a heroic society is archaic Greek culture as presented in the *Iliad*. (Another example of his is the mediaeval Norse culture of the Icelandic sagas.) According to MacIntyre, members of such heroic societies have all-encompassing roles such that they cannot intelligibly step back from these roles and ask why or whether to act according to such roles. One such role is that of the *war-hero*. Such roles have 'constitutive standards.' These are standards from which an individual may deviate to some degree, but still fall under the category. In that case, we can call them a *bad* or *imperfect* instance of the category. A warlord who is *sometimes* cowardly may still be a war-hero, but he would be an imperfect one.

As MacIntyre sees it, Achilles cannot *deliberate* or *reason* on a question like: 'I am a war-hero and a war-hero makes war, but still: why make war?'. Of course, Achilles may weigh his reason to make war in some particular circumstances against other reasons a war-hero has. If, say, going to war would show disregard for his own honour after an insult, perhaps he oughtn't make war. However, if it is settled how a perfect war-hero would act in the situation, then there is no further room for reasoning or deliberation. Thus, according to MacIntyre, the role of a war-hero is beyond ethical criticism for Achilles. There are no other, conflicting roles against which Achilles could weigh the role of a war-hero. All other normatively relevant roles, such as that of a Greek man and noble, are subsumed under that of a war-hero.

Furthermore, asking whether it is *good* or *right* to act like a war-hero is otiose. If the question even makes sense, the answer is *obviously* 'yes'. He cannot let reasoning intervene between recognising the requirements of the role, and action. Achilles, of course, is torn between continuing to make war against Troy, or to go home. However, for Achilles and his social environment, this conflict is not amenable to reasoning. Any mental stirring in this vicinity can only be understood as weakness of will (or outright madness).

The *poet* writing about Achilles may ask who lives the better life: a heroic warlord, or a peaceful ruler at home. However, this poet does not live in the same society as Achilles. (As a matter of actual timelines, the Iliad's orally transmitted text was fixed in the 8th century BCE, while the events originally inspiring it took place around the 12th century BCE.) For Achilles himself, so MacIntyre,

this is not (yet) a question about which one could deliberate. In Adorno's terminology, for Achilles and his contemporaries, the role of the war-hero was a piece of "myth": beyond change or critique, and without a need for justification.

In contrast, in modern forms of ethical discourse, agents can weigh any requirement entailed by one role against requirements entailed by other roles. They can also ask if a given role is right or good. Function-based naturalism can vindicate *this* fact about it. However, concepts of roles such as that of a war-hero are not the only concepts we have of *ways of being human*. Evolutionary naturalism cannot explain how we may step away from some other ways of being human. The only forms of discourse that could be vindicated by such function-based naturalism are ones which treat the notion of a *well-functioning human being* in the same way as the ethical discourse of heroic societies treats roles like *war-hero*. In the forms of ethical discourse vindicated by function-based naturalism, asking the following would be otiose: 'Sure, this is how a well-functioning human being would act, but why act that way?' As far as function-based naturalism is concerned, any reasoning intervening between the recognition of what a well-functioning human would do and acting that way would be mere superstition. One may try to vindicate such reasoning another way, e.g. through a non-cognitivist or constructivist account. However, in that case, function-based naturalism would not explain the intimate link between ethical discourse and action. Rather, it would be the form of discourse captured by the latter account that has this intimate link. Thus, function-based naturalism would not really be an account of ethical discourse.

Thus, whether function-based naturalism can vindicate ethical discourse depends on the notion of a *well-functioning human*. The function-based naturalist must show that this is not the concept of a (contingent) way of being human which we can subject to ethical criticism. Failing that, they need to show at least that our ethical discourse will not be altered unacceptably if we treat it that way. That is, it allows the same kind of ethical critique of particular prescriptions, only with some redescription. Otherwise, function-based naturalism fails to vindicate ethical discourse. It would merely vindicate a hypothetical form of ethical discourse which is not ours. Thus, function-based naturalism would fail because it would hypostatise a certain way of being human.

Given function-based naturalism, this worry can also be described as a worry about hypostatising particular *circumstances* of human life. Suppose it is biologically typical for individuals of a species in a

given environment to develop in a certain way. (By *biologically typical*, I mean that most individuals which contribute to the continuation of the species develop that way. Most mayflies die before they can reproduce. Yet, it would not be biologically typical for them to die so early.) If it is biologically typical for individuals of a species to develop in a certain way in a given environment, then either the way they develop is vitally functional, or the environment is *unfit* for them. Thus, if an account of ethical discourse hypostatises a certain way of being human, then it also hypostatises the environment where human beings typically develop to be that way, or in which they have to be that way in order to survive. It makes it pointless to ask whether such an environment is fit for human beings.

I will illustrate the danger of hypostatisation by considering two versions of evolutionary naturalism. Both provide an account of the *well-functioning human being* which we can step back from in modern moral discourse. The two versions are *propensity-based naturalism* and *etiological naturalism*. According to propensity-based accounts, the function of a trait consists in the propensity it has to contribute to (inclusive) fitness, i.e. to the unit of selection being reproduced. Thus, it is irrelevant whether or how a trait has been selected. Given that it has a propensity to contribute to exclusive fitness, it has a function. Conversely, if it lacks such a propensity, it does not have a function. The function of my heart is to pump blood because by pumping blood it increases the chances that my genes be passed on. It corresponds to accounts of function proposed by Bigelow and Pargetter (1987), Kitcher (1993) and Boorse (2014).

Etiological evolutionary naturalism, in contrast, is based on accounts which explain biological function through their selection history. Such accounts have been defended e.g. by Millikan (1989) and Garson (2022). According to such accounts, broadly, a trait T has the function F if it is a copy of traits T* which were only copied because they F'd. My heart has the function to pump blood because it is the copy of other hearts which were only copied because they pumped blood. Thus, according to such accounts, my faculty of practical reason is the copy of my ancestor's faculty of reason (and that of their kin who contributed to their reproductive success). These 'originals' were copied only because they had a certain effect. Performing the same effect is the function of my practical reason. I am practically rational (and virtuous) insofar as my practical reason is apt to perform that effect.

3.1 Propensity-based naturalism: Hypostatising fitness-maximisation and the 'Struggle for Existence'

Propensity-based naturalism hypostatises fitness-maximisation as a way of being human. It cannot explain how human agents may step away from being a *fitness-maximiser* as a way of being human. The concept of a *fitness-maximiser* has constitutive standards. An agent may sometimes fail to behave in ways that maximise fitness and be an imperfect fitness-maximiser. However, if they deviate too much, they are no fitness-maximier at all. (One might question whether anything in the world corresponds to this concept, such that the standards are grounded in what it is, rather than being arbitrary projections of the assessor's mind. Korsgaard assumes this is the case with constitutive standards. We will set this objection aside for argument's sake.)

In any form of ethical discourse which propensity-based naturalism can vindicate, the question 'this is what a (perfect) fitness-maximiser would do, but why act like a fitness-maximiser right now?' would be pointless. Consequently, it also hypostatises the circumstances under which it is statistically typical for human beings to act as (approximate) fitness-maximisers. These are circumstances under which human beings must act as fitness-maximisers in order to survive or attain other goals they have. These circumstances can be called the *struggle for existence*, using a phrase associated with Malthus (2008 [1789]) and employed to great effect by Darwin (1910 [1859]). With regard to human beings, these describe circumstances that are characteristic of Victorian capitalism, or perhaps capitalism in general.

The American oil tycoon J.D. Rockefeller (1839-1937) provides a vivid illustration. He told a Sunday school class (!): "The growth of large business is merely a survival of the fittest. [...] This is not an evil tendency in business. It is merely the working out of *a law of nature* and *a law of God*" (see Lux 1990: 148, my emphasis).

The consequences of this hypostatisation can be seen from the particular ethical commitments it leads its exponents to. In all of these cases, the considerations presented by the defenders may show that the behaviour they argue is virtuous is appropriate for humans living under certain circumstances, but they do not show that it is good for humans as such. Ragnar Redbeard (1910 [1896], apparently the pen name of a cattle-drover and political activist from New Zealand called

Arthur Desmond, see Parker 1982) endorses a thorough-going egoism. Leo Tolstoy (1899: 159) describes Redbeard's *Might is Right or The Survival of the Fittest* as such:

> The substance of this book, as it is expressed in the editor's preface, is that to measure "right" by the false philosophy of the Hebrew prophets and "weepful" Messiahs is madness. Right is not the offspring of doctrine, but of power. All laws, commandments, or doctrines as to not doing to another what you do not wish done to you, have no inherent authority whatever, but receive it only from the club, the gallows, and the sword. A man truly free is under no obligation to obey any injunction, human or divine. Obedience is the sign of the degenerate. Disobedience is the stamp of the hero. Men should not be bound by moral rules invented by their foes. The whole world is a slippery battlefield. Ideal justice demands that the vanquished should be exploited, emasculated, and scorned. The free and brave may seize the world. And, therefore, there should be eternal war for life, for land, for love, for women, for power, and for gold. (Something similar was said a few years ago by the celebrated and refined academician, Vogüé.) The earth and its treasures is "booty for the bold."

The book is rife with contradictions, stemming from his endorsement of normative egoism alongside white supremacy, racial anti-Semitism and the patriarchal family, which all require identification with a collective. While not at all well-known nowadays, the pamphlet elicited Tolstoy's above-quoted response as well as a discussion by Paul Carus (1899) in *The Monist*.

Most propensity-based naturalists in contrast endorse arguments in a Hobbesian style. A monomaniacal focus on narrowly construed self-interest or on intentionally maximising one's inclusive fitness does not in fact maximise fitness. This is firstly because conscious human reasoning is both costly and prone to errors. For example, we may be prone to overestimate our ability to avoid being detected in rule-breaking behaviour. Thus, an agent who treats a rule like 'don't lie and cheat' as providing (defeasible) non-instrumental reasons may be more fit than one who treats it only as an instrumental counsel. Secondly, even abstracting away from particular cognitive limitations, it can be an evolutionary advantage for an agent to commit to a future action. For this, an agent needs to treat it as a reason for an action that they have committed to such an action in the past, even if the action

does not now maximise their self-interest or inclusive fitness. Such an agent can, for example, assure or promise to conspecifics that they will act in a certain way. This enables forms of cooperation which, over the long run, are beneficial to the agent (with regard to self-interest and fitness).

An individual propensity-based naturalism informed by such considerations has been propounded by Herbert Spencer (1879). Like Redbeard, he appeals to the "survival of the fittest" as a normative principle applying both to humans and in the rest of living nature. However, he believes that it implies the "Principle of Equal Freedom". According to it, "each has freedom to do all that he wills provided that he infringes not the equal freedom of any other". He even recognises the uses of charity in maintaining stable patterns of cooperation. However, he is rightly notorious for the withdrawal of altruism from the weak and destitute:

> "Besides an habitual neglect of the fact that the quality of a society is physically lowered by the artificial preservation of its feeblest members, there is an habitual neglect of the fact that the quality of a society is lowered morally and intellectually, by the artificial preservation of those who are least able to take care of themselves. . . . For if the unworthy are helped to increase, by shielding them from that mortality which their unworthiness would naturally entail, the effect is to produce, generation after generation, a greater unworthiness." (cited in Casebeer 2003: 94).

Helping the poor and "feebl[e]" is unlikely to benefit your fitness. They won't be able to help you in reciprocal cooperation. They are similarly unlikely to contribute to the "quality" of your "society" and thus to the likelihood that the genes you share with the fellow members of your society will be reproduced through its collective success. However, in our actual ethical discourse, we are able to ask 'a fitness-maximiser would be indifferent towards the most needy and vulnerable, but still: why act with indifference'. 'Vulgar' propensity-based naturalism cannot explain this aspect of ethical discourse. Similarly, it cannot explain how we can ask whether an environment in which human beings develop to be (approximately) fitness-maximisers is unfit for human beings. Thus, it fails to vindicate modern ethical discourse.

This objection cannot be avoided by focussing on groups of individuals as the unit whose fitness is under consideration. Group-focussed accounts of propensity-function have not been defended

explicitly in the philosophy of biology. However, many biologists and philosophers of biology have argued for the importance of the social group as a unit of selection, e.g. Okasha (2006). A collective propensity-based account has been defended by Arthur Twining Hadley. The economist, railway expert and 13th President of Yale University hybridises Spencer's evolutionary naturalism with American pragmatism in his *Some Influences in Modern Philosophic Thought* (1913: 35). He endorses the following definition of *right*:

> The scientific man … is concerned to study the right and wrong of things; and he believes, as the very essence of his theory, that the right is that which will prevail in the long run. The criterion which shows whether a thing is right or wrong is its permanence. Survival is not merely the characteristic of right; it is the test of right.

Strictly speaking, he does not identify the rightness of a "thing" with its propensity to contribute to its survival, but with its actual survival. However, the influence of Darwin is evident, e.g. in the appeal to the "scientific man" and to "survival". A principle or practice is right if it itself survives. That is, it is right, for Hadley, if it is part of the worldview of a group which passes this worldview on to their descendants and which secures its own survival, including in the conflict with other groups. Even more than Spencer, Hadley endorses so-called Social Darwinist ideas. He argues that it is conducive to a culture's survival, and thus right, for it to prevent its 'weakest' and 'least fit' members to reproduce. He is of special interest because of his influence on Aldo Leopold, who continues to be relevant to green thought and environmental ethics.

William A. Rottschaefer and David Martinsen (1990) present a much later defense of collective propensity-based naturalism. They are primarily concerned to argue against Michael Ruse that morality is "objective", i.e. that moral terms denote real, mind-independent properties. To that end they identify moral goodness with the propensity of an action to advance "human fitness". "Human fitness", they mean in the sense 'fitness of the human species' (rather than, say, 'characteristically human fitness). However, they clarify that "[i]t is not clear to us that the evolutionarily based value of human fitness includes all humans" (1990) rather than "a person's family and rather immediate relatives [or something] between the extremes" (1990, n. 8)

However, such accounts do not escape the worry about hypostatisation. 'This is what a group-fitness-maximiser would do, but still: why do it?' is an open question. Furthermore, the view has similar implications concerning charity towards the most needy. Their survival, plausibly, is not conducive towards the survival of the group. It also hypostatises the struggle for existence. If the relevant group is smaller than the nation, it hypostatises the struggle between such groups. And even if the group encompasses the whole species (i.e. humanity), it still hypostatises humanity's 'struggle for existence' as its transhistorical conditions. It excludes the possibility that circumstances in which humanity freely shapes its own future may be more fit for humans.

3.2 Etiological naturalism: Hypostatising (all-too-)human nature

Etiological naturalism does not hypostatise fitness-maximisation or the 'struggle for existence'. According to etiological naturalism, it is possible for a trait to contribute to fitness and yet not be functional. For example, it may be that competitiveness contributes to fitness in modern societies, but that it has not done so until very recently in human's evolutionary history. Thus, plausibly, competitiveness would not have been selected for, so it would not be a functional trait. Hence, it makes sense to ask 'this is how a fitness-maximiser would act, but still: why act that way?'

However, etiological naturalism hypostatises what we might call the way of being human as a 'species-typical *Homo sapiens*'. This use of *species-typical* is based on Neander (2017). A species-typical individual is one which does not have any dysfunctions, i.e. one in which every trait realises the effects it has been selected for. 'Species-typical *H. sapiens'* has constitutive standards. An individual who falls short of these standards to some degree may still be an '**imperfectly** species-typical *H. sapiens*'. However, if they fall short by too much, they are not a species-typical *H. sapiens* at all. According to etiological naturalism, it should not make sense to ask 'this is how a species-typical *H. sapiens* would act, but still: why act that way?' In modern ethical discourse, however, this makes sense. This is, somewhat surprisingly, illustrated by Nietzsche (2016 [1878]) . In §224 of *Human, All-Too-Human*, he criticises Darwinists[5] for not seeing that "degeneracy" (*Ent**art**ung*), or deviation from the species (*Art*), may contribute to the "ennoblement" (*Veredelung*) of the individual and her kind. Human beings, for Nietzsche, ought to overcome the way their nature is defined by their past rather than persist in it.

[5] He names "Darwin" as his target, though it appears that he was mostly familiar with Darwinism through Ernst Haeckel (e.g. 1899). Unlike Darwin, Haeckel defended evolutionary naturalism in ethics.

Consequently, etiological naturalism also hypostatises the environment in which human beings evolved. It is otiose to ask whether such an environment is fit for human beings. After all, this is the environment for which they have been adapted. In this environment, their traits are most likely to contribute to fitness in the way they have been selected for. Furthermore, the capacities for development are most likely to unfold in (one of) the way(s) which they have been selected for, thus resulting in species-typical rather than abnormal traits.

At first, this might seem to be the correct result. After all, which environment could be more appropriate for a species than that in which it evolved? However, it is not adequate to the role which the notion of an environment fit for human beings plays in our ethical discourse. Virtually for the entirety of their evolutionary history, for instance, modern humans lived without states, without the written word and without agriculture. The presence of these things almost certainly changes their environment such that some selected-for mental traits hinder rather than produce fitness, and some developmental capacities 'misfire', leading to humans developing dysfunctional traits. However, it does not follow that an environment with these things is unfit for human beings. In fact, one may intelligibly hold that an environment that contains such things as written poetry and representative government is *more* fit for human beings. Such a position may be false. However it is not *obviously* so, even keeping the evolutionary facts in mind.

This form of hypostatisation can be illustrated further with three particular examples. Two of these concern behaviours which we would clearly judge reprehensible. *A fortiori*, they are subject to ethical critique. These are racial in-group favouritism and sexual violence. The last concerns apparently biologically based gender roles. While following such roles is not *per se* reprehensible, it is an important feature of our ethical discourse that we can step away from them.

The first example concerns racial in-group favouritism. Various philosophers and psychologists have argued that dispositions or 'psychological devices' for racial categorisation are an evolved aspect of the human mind (e.g. Gelman 2003, Machery et al. 2010, Kelly et al. 2010, Mallon 2013). By racial categorisations, I mean categorisations of people based on easily observable phenotypic features (like skin colour, hair type, height, facial features) which are taken to imply important facts about their bearer. Many researchers also include one or two further aspects in defining race-thinking:

racial categories must be assumed to be heritable, and there must be a hierarchy between them. These features are the most controversial in the cross-cultural evidence. They are not relevant to my argument.

Such racial categorisations benefit the reproductive fitness of the individuals making them. Communities that distribute social roles in part based on obvious phenotypic features of their members are highly legible. They make it easy for their members to quickly comprehend a social situation, including one involving strangers. For individuals living in such a society, it is advantageous to quickly categorise people based on the relevant phenotypic features. It is also beneficial to take such categories to imply important further facts about the bearer. After all, they do, even if those facts are socially constructed.

So far, this would pose no problem for etiological naturalism. Noticing race is not a vice. In fact, so-called 'colour-blindness' is often criticised as an ideology that hides continuing white supremacy. However, there is also evidence that people favour members of any group they consider themselves to belong to. They are less likely to criticise them harshly, more likely to cooperate with them and more likely to help them.

Such a tendency may well have evolved to enable a kind of indirect reciprocal altruism. Smaller subgroups have a better chance at maintaining a pattern of mutual support and exclude free-riders than a larger population (Efferson et al. 2008; Fu et al. 2012). If the group membership in question is heritable, in-group favouritism also contributes to a more favourable social environment for one's offspring.

Combining such a tendency with the thesis that racial categorisation is selected for, it follows that it is in line with the evolved functions of their mental dispositions for humans to favour conspecifics who share their contextually relevant phenotypic traits, i.e. who are of the same race. This applies inter alia to members of dominant racial groups in racist societies. Thus racist biases in hiring, granting credit, policing and the justice system would be in line with the evolved functions of the dispositions in the racist's mind. Thus, etiological naturalism would be committed to seeing such racial in-group biases maintaining racial hierarchies as virtuous. They cannot explain how there is a

point to asking: 'Siding with my own race under these circumstances is how a species-typical *H. sapiens* would act, but still: why act that way?'

The second example of such behaviour concerns sexual violence perpetrated by men against women. While rates of incidence vary, vaginal-penile intercourse forced through violence or the threat of violence has been reported from nearly all cultures. While it appears that only a minority of men in most societies are rapists, rates of incidence are shockingly high. Studies find that up 13% of women in the US have experienced rape, and sexual violence is likely under-reported. Rape makes an obvious contribution to fitness. Absent modern methods of contraception, it allows the rapist to have more offspring. While it would not have contributed to the victimised women's inclusive fitness to have chosen the rapist as a mate, if they are impregnated by him, throughout evolutionary history, they would only have been able to abort the pregnancy at a high risk to their health and future fertility. Thus, it often increases their inclusive fitness to carry the child to term and raise it, since it also carries her genes. Hence insofar as the behaviour in question has been shaped by natural selection, evolutionary theory predicts that they are disposed to do so. It also predicts that males are disposed to rape females if they can do so with a low risk of punishment. Indeed, forced copulation is not only common among humans, but among other primates. Evolutionary psychologists have developed accounts of rape which explain particular empirical observations (McKibbin et al. 2008) and developing means to prevent rape and counteract rape culture (Vandermassen 2011).

The perpetration of sexual violence involves treating considerations as reasons for action. Rapists treat the vulnerability of the victim and their opportunity to exercise power over them as reason to act. Thus, according to the hypothesis, the disposition to perpetrate sexual violence would be a selected-for function of practical reason. Furthermore, it likely still contributes to inclusive fitness. Thus, the etiological naturalists would be committed to seeing sexual violence in certain circumstances as ethical. They cannot explain how there is a point to asking: 'Perpetrating sexual violence is how a (male) species-typical *H. sapiens* would act in these circumstances, but still: why act that way?'

A final example concerns gender roles. It is likely that some behavioural differences between male and female humans have been selected for. One example is parental involvement with child care. While there is very significant variation between cultures, in most societies, men spend less time with

their young children and invest less effort in child care. This holds true even after weaning. It can be explained through evolutionary considerations. Paternity is less certain than maternity. Thus, game-theoretically, fathers 'should' invest less resources into any individual offspring. It is also less costly for a male to conceive another child. He does not need to go through pregnancy. Thus, it is likely that some element of this behaviour is due to selected-for dispositions. It is also plausible that there are other such sexually dimorphic dispositions.

Nonetheless, in modern ethical discourse, it is possible to ask 'this is how a species-typical female/male *H. sapiens* would act, but still: why act that way?'. The cogency of these normative standards to one's life can be questioned in at least two ways. Firstly, one may question whether (or to what extent) to act by the particular content of gender roles implied by the biological inheritance of *H. sapiens*. That is, I may accept in general that norms for (say) female human beings apply to me, but reject some particular such norms. I may instead judge that males and females ought to be equally involved in child rearing. Secondly, one may question one's own 'gender identity'. That is, I may intelligibly deliberate about whether the norms for 'male human beings' apply to me even if I accept that I am 'biologically male'. Such reasoning may, for example, take the shape of seeking an interpretation of one's inner life and/or one's life-history so far.

3.3 Why not embrace biologism?

Evolutionary naturalists may argue that it is appropriate to place certain ways of being human beyond ethical criticism because they are inevitable. Male human beings, they may say, cannot but be disposed to commit sexual violence in certain circumstances. By a principle of 'ought implies can', it is inappropriate to critique such dispositions. Noting this is no more an objectionable hypostatisation than to note that human beings, by nature, cannot fly.

However, evolutionary theory itself implies that there is some variation. Without variation, there could be no evolution through natural selection. Furthermore, experience and ethnographic evidence suggests that variation in human behaviour is much larger than could be explained by genetic variation alone. Rather, human behaviour develops through the interaction of genes and environment. Only some of these developmental trajectories have been selected for. The etiological naturalist claims that the selected-for trajectories are also the only ethically good ones. This is distinct from the question concerning the extent of possible human variation.

It may turn out that some disposition towards, say, sexual violence is inevitable for human beings. If so, it may well follow that --- analogously perhaps to the disposition to fall asleep at certain intervals --- this disposition cannot be criticised ethically. Ethical critique plausibly should instead concern itself with ways to manage this disposition and keep it from being actualised. However, it does *not* follow that the (internal and external) conditions under which this disposition will be actualised as selected for are beyond ethical critique.

A defender of evolutionary naturalism may also argue that we should revise the form of our ethical discourse so that it is in line with a form of evolutionary naturalism. After all, if MacIntyre is right, it is possible for human beings to live with a similar form of ethical discourse. What grounds do we have to assume that modern ethical discourse is superior or more fit for human beings? Just like we shouldn't *a priori* expect that a naturalistic account of ethics vindicates all of our ethical judgements content-wise, so we shouldn't expect it to vindicate the precise form of our ethical discourse as it is. One can easily see, say, Redbeard arguing that any reasoning that intervenes between recognising what is necessary to succeed in the 'struggle for existence' and so acting constitutes weakness and defect.

However, firstly, MacIntyre might not be correct in his reading of *Iliad*. Sophie Grace Chappell (2023: 33) argues that characters in the epic can step away from their roles and question the reasons provided by these roles. Consider Achilles' thoughts while he is torn between re-entering the fight against the Trojans and going home:

> My mother, silver-footed goddess Thetis,
> foretells that I confront now double fates,
> a choice of roads before me to life's end.
> The one road is: remain and fight at Troy,
> never get home, but win eternal fame.
> By the other road, by return to my dear Phthia,
> "Your glory is extinguished, but Achilles,
> you'll live long years at home, safe and obscure,
> evade the doom of death, and grey endure."

If MacIntyre is right, the last three lines describe an outcome that may be *tempting* to Achilles. However, for him there is no question that he has no (all-things-considered) *reasons* to choose it. For Chappell, however, Achilles is genuinely considering whether to give up the role of war-hero or continue in it. Thus, contra MacIntyre, he demonstrates an ability to detach from his role. Chappell (2023: 44) suggests that "Homeric society is about as different from our society as any human society could possibly be". Thus, any human society allows for role-recalcitrance.

I am not quite sure if this last assumption is warranted. The Greeks of the communities described by Homer are indeed much further from us than even those of the classical period. However, they live off trade, farming and fishing. Their religion is focussed on a small number of immensely powerful, personal gods. It seems that, say, a foraging society whose religious practices are centered on a more diffuse 'spirit world' may be much further from us while still being recognisably human. Be that as it may, if Chappell's interpretation is correct, this invalidates an important piece of evidence that a 'heroic' form of ethical discourse is possible for human beings.

Furthermore, the conceptions of ways to live a human life which evolutionary naturalism hypostatises differ significantly from the conception of roles considered by MacIntyre. Roles like *war-hero* entail purposes that are external to their bearer. The war-hero serves his tribe or nation as a protector. This does not, of course, mean that Achilles is motivated by 'altruism'. Most of all, he appears to seek glory or fame. However, glory must be recognised and celebrated by a community to be achieved (cf. Chappell 2011). In this, MacIntyrean roles resemble concepts of tools, which also have external purposes. In contrast, concepts like *fitness-maximiser* or *species-typical H. sapiens* do not imply such an external purpose. Thus, it may be that, while it is possible for human beings to live within a form of ethical discourse where they cannot step back from such roles, they can always step back from the latter ways of being human.

Last but not least, just because human beings are in principle capable of living within such a form of ethical discourse, it does not follow that modern ethical discourse can be reformed to be that way. To us, it at least seems intelligible for reasoning to intervene between the recognition of what a well-functioning human being would do and action. If we were to attempt to live within an ethical discourse that doesn't allow such reasoning, one of two things would happen.

On the one hand, we might attempt to *suppress* reasoning between the recognition of what a well-functioning human being would do, and action. However, the resulting form of ethical discourse would not be like MacIntyre's heroic society. It would be one in which such reasoning is intelligible, and in principle possible, but simply suppressed. This is not a form of ethical discourse which could be vindicated by evolutionary naturalism.

On the other hand, one may try to use the arguments supporting evolutionary naturalism to 'bridge the gap' between the recognition of what a well-functioning human being would do and action. An agent finding themselves asking why to act like a well-functioning human being may remind themselves, say, of how such reasoning is weak and defective, or how it leads to corrosive 'historicism' and 'relativism' (cf. Strauss 1953, who influenced Wilson 1993 and Arnhart 1998).

However, such a mental activity would constitute exactly the kind of reasoning which must be otiose according to evolutionary naturalism. Furthermore, the arguments the agent reminds themselves of either amount to arguments that a well-functioning human being would not engage in, or they present some other disadvantage of such reasoning. In the first case, they simply beg the question. There will be room for reasoning to intervene between *these* arguments and action. In the second case, they are incompatible with evolutionary naturalism.

4. Conclusion

In this chapter, I have introduced the notion of function-based naturalism. I have argued that function-based naturalism faces an apparent dilemma between meeting the challenge from science and avoiding hypostatisation. I have also argued that the forms of biological naturalisms found in the literature, which are based on evolutionary theory, fail to resolve the dilemma. They hypostatise certain ways of being human, i.e. they place them beyond ethical criticism. In the next chapter, I will focus on the neo-Aristotelian attempt to resolve the dilemma.

THREE: Neo-Aristotelian naturalism and its discontents

In the last chapter, I have introduced function-based naturalism and the dilemma between responding to the challenge from science and avoiding hypostatisation. In this chapter, I will consider the neo-Aristotelian response to this challenge. I will first introduce that response. Then, I will consider different arguments for the claim that neo-Aristotelianism can respond to the challenge from science. I argue that all of these arguments fail.

1. Neo-Aristotelian naturalism

Neo-Aristotelian naturalism attempts to avoid the 'threat of "biologism"' by denying that judgements of vital normativity can be reduced to judgements belonging to some other discourse, including (other) parts of the life sciences. One cannot give an account of the nature of vital normativity which would make it possible to derive judgements of vital normativity from judgements of empirical biology. Thus, the " 'autonomy of ethics' " (Thompson 2004) is preserved. Ethical judgements cannot be derived from empirical, non-ethical descriptions of human life. Consequently, no substantial way of being human is removed from ethical criticism.

For the neo-Aristotelian naturalist, one cannot sensibly ask, say: 'sure, the naturally excellent human being is competitive, but still: why act competitively?' However, one can always intelligibly raise questions concerning whether competitiveness really is a vital function of human practical reason, i.e. whether the naturally excellent human being would really be competitive. Furthermore, since competitiveness is now understood as a trait of the naturally excellent human being, one can always raise questions about what it really consists in. Compare how we can always ask whether a courageous person would really choose to face a certain danger, or whether doing so is actually reckless.

Furthermore, the reason to judge that, say, competitiveness is a vital function of human practical reason is also *ipso facto* a reason to treat some action's expressing competitiveness as a reason to act. Thus, Korsgaard's/Street's 'normative question' objection could be avoided. Whenever an agent is treating a consideration as a reason for action, there is also some consideration which they are (at least implicitly) treating as a reason for doing so. If Street is right, they must, of course, also have a

reason for treating this second, 'higher-order' reason as a reason. However, this does not necessarily lead to a vicious regress. Rather, there is a 'web of reasons', such that every treating-as-a-reason is supported by some other treating-as-a-reason (cf. Street 2012).

While other neo-Aristotelians allude to such ideas, Michael Thompson develops this response in most detail (see Thompson 2004, 2013). He argues that our ways of thinking about living beings depend on special *forms of judgements*, i.e. on judgements which differ from judgements about, say, mathematical objects or lifeless physical things not just with respect to the concepts featured in them, but with respect to the way in which these concepts are combined and how judgements relate to other judgements. Specifically, *vital judgement* --- judgements about particular living beings as living beings, say that *my dog Fido is eating* --- imply *natural-historical judgements*. These are judgements about the life-form the particular living being bears. They can be naturally expressed with generics about a species or life-form, e.g.: 'Gulls can fly' or 'the dog has four legs'. (I will use the singular with a capitalised noun for the life-form in order to explicitly designate Thompsonian natural-historical judgements: *The Dog has four legs*.) The true natural-historical judgements about a life-form describe its natural history, i.e. roughly the 'species" or life-form's characteristic way of life. According to Foot (2001), a trait is a "natural excellence" (i.e. vitally functional) if individuals of a life-form need it in order to achieve some good according to their natural history.

Thompson suggests two reasons why natural-historical judgements are not reducible to other judgements in empirical biology. Firstly, they draw on the totality of a subject's knowledge about a kind of living being, and they do so in an interpretative way. In this, they resemble aesthetic judgements, as when I judge that some music is solemn, melancholic or joyful. While I can point to specific features of my observation of particular individuals of a species to support my judgement that a certain behaviour belongs to its natural history and another behaviour doesn't, this process could never be described exhaustively through rules. For any partial description of the kind of living thing which supports a certain natural-historical judgement, there could be a context in which it does not support that judgement.

Secondly, we *always already* think of living beings through vital judgements and natural-historical judgements. There is no 'given' in biology to which vital judgements could be reduced. If a biologist encounters, say, a new species of jellyfish, they will at first try and subsume it under a life-form they

are already familiar with. They will treat it as an aberrant specimen of a known jellyfish-species. Otherwise, so Thompson, they could not make vital judgements about it and thus not recognise it as a living being at all. As they observe the jellyfish and encounter more individuals of its kind, the biologist will revise their judgement and consider them as a separate life-form. However, their observations depended on the early vital judgements and the natural-historical judgements they implied.

An objector might wonder whether neo-Aristotelian naturalism even deserves the title 'naturalism'. After all, according to it, it might turn out that the vital function of practical reason is to disclose some seemingly non-natural property, e.g. *goodness*, *rightness* or *being a reason*. The accounts of, say, Hursthouse (1999) and McDowell (1998) may be read that way. However, if it really turns out that the vital function of practical reason is to disclose rightness, then rightness features in at least one judgement that is continuous with the natural sciences: precisely the judgement that disclosing rightness is the function of practical reason. Thus, rightness can be called a natural property. This is so even if it cannot be reduced to other natural properties. Such forms of neo-Aristotelian naturalism thus have a more expansive notion of 'nature' than is presupposed in the objection.

In order for the neo-Aristotelian response to be successful, the neo-Aristotelians must show (a) that their response avoids the 'threat of "biologism"' and especially that it does not hypostatise contingent ways of being human; and (b) that it does so without falling prey to the challenge from science. That is, it must show that the subject matter of ethics is continuous with biology such that one cannot coherently deny the legitimacy or scientific respectability of ethics without also denying the legitimacy of much of biology. In the following, I consider various arguments presented by neo-Aristotelians for (b). I argue that none of these arguments succeeds. Instead, a *robust naturalism* is needed if the naturalist response to the challenge from science is to succeed, i.e a view which shows ethical judgements to be continuous with judgements made *within* natural sciences.

2. Vital judgements as the 'manifest image' of life

Thompson (2008) repeatedly describes vital judgements as constituting the "manifest image" of living nature. The notion of the *manifest image* is taken from Wilfred Sellars (1962). According to

Sellars, we have two modes of describing the world: the *manifest image* and the *scientific image*. The manifest image is the conceptual scheme in which human beings first encountered themselves and their environment. It contains concepts of rational agents and their actions as well as of the objects they are acting on: 'mid-size dry goods' along with similar ordinary physical objects, and other living beings. The manifest image is concerned with *capturing* regularities in the environment of human beings but not with explaining them. In contrast, the *scientific image* explains such regularities by postulating unobservable entities that cannot be directly acted on by human beings. This distinction, one might think, is enough to show that natural-historical judgements are not superstitious or pseudo-scientific. Just as our 'manifest image' judgements about non-living material objects have not become superstitious through the discoveries of modern physics and chemistry, so our 'manifest image' judgements about living nature remain scientifically respectable despite modern biological discoveries.

However, Thompson's appeal to the 'manifest image' alone cannot dispel the challenge from science. Firstly, there are important differences between the relation of everyday judgements to biology and to physics or chemistry. The latter fields, however, are what Sellars had in mind when he developed the distinction between manifest and scientific images. Unlike physics or chemistry, evolutionary biology does not need to make reference to non-observable entities. While evolutionary biology can be stated in ways that refer to, say, abstract *genes* or *DNA molecules* too tiny to be observed by ordinary means, much of evolutionary biology can be described simply by referring to observable characteristics of individual animals along with the heredity of those characteristics. This is how Darwin presented his view in the *Origins of the Species*. While, say, Newton's *Principia Mathematica* describes a mathematical system, constituted by axioms and populated by unobservable entities like forces or momentums, and by bodies large enough that human beings cannot experience or act on them as bodies, Darwin's (1910 [1859]) *Origin of the Species* builds up its proposal by generalizing from human activities like animal husbandry and pigeon-fancying.

Hence, evolutionary biology makes statements that are straightforwardly incompatible with platitudes implied by the manifest image of living nature according to Thompson. One 'manifest image' platitude is this: Through reproduction, an individual can only produce fertile offspring belonging to the same species. The inference 'this animal is an [elephant, piranha, goat] so its parents must have been [elephants, piranhas, goats]' is characteristic of reasoning about animals and acquired

by children as young as three years old (Gelman 2003). Our 'manifest image' understanding of life also contains the concept of *ancestor*: 'a parent, or a parent of a parent, or …' for any finite iteration of 'a parent of'. By simple logical principles, it follows that all ancestors of, say, an elephant have also been elephants, etc. This is incompatible with evolutionary theory, which holds that different species share common ancestors.

There is no equivalent direct incompatibility between the manifest image and physics or chemistry. Similar-seeming paradoxes can be generated using indefinite iterations. However, in all of these cases, the paradoxes depend on concepts of entities falling well outside the domain of the manifest image. In contrast, with evolutionary biology, it is only the *relation* of being a (say) 500,000×-great-grandparent[6] that falls outside the manifest image. For example, according to the 'manifest image' conception of a stuff, every part of a piece of emerald is itself a smaller piece of emerald. According to chemistry, some parts of emeralds, in contrast, are beryllium atoms. However, the diameter of a beryllium atom is at most 0.5 nanometres: 1/2,000,000 of a millimetre or 1/200,000 of a human hair, i.e. of approximately the smallest breadth perceivable with the unaided eye. The concept of such a small bit of matter has no place in the manifest image. It is not the kind of object human agents might directly act upon. Similar points can be made for other apparent paradoxes. For instance, in the manifest image, every event is preceded by some other event. In the scientific image, the Big Bang is not preceded by anything. However the notion of the Big Bang has no place in the manifest image. In order to construct the contradictions between physics or chemistry and the manifest image, one must iterate in order to construct concepts of entities that do not belong to the manifest image. If one were to split an already tiny piece of emerald 200,000 times, the procedure would pretty soon be unobservable to human beings. In contrast, each of the about 500,000 births connecting me to the latest ancestor I share with a chimp was in principle capable of being observed by a human being.

Philippa Foot (2001: 29) suggests that natural-historical judgements describe a "stil[l]" of evolutionary history. However, this metaphor is misleading. The processes constituting a life-form's natural history are extended in time. This is true of growth and nutrition as much as of reproduction. Foot's remark encourages thinking of this time-dimension as being 'orthogonal' to the

[6] This is a very rough estimate for the last common ancestor of modern humans and modern chimps. This ancestor lived some time between five and thirteen million years ago.

time of evolutionary change. Evolutionary history is a succession of different life-forms. Each life-form has its own natural history. As far as evolutionary history is concerned, this natural history would be contained in a single instant. Yet, evolutionary history takes place in the same time-dimension as the life-cycle of individual organisms. It simply occurs on a much larger time-*scale*. Mutation, i.e. the slight change of hereditary characteristics enabling evolution, happens during individuals' reproduction.

Here is one seemingly straightforward way to reconcile the two perspectives. From the point of view of natural history, mutations are mere accidents. Some of them are defects, some merely neutral abnormalities. However, over the time-scale of evolutionary history, these small mutations add up such that new species or life-forms emerge.

However, first of all, while conservative, this reconciliation requires an adjustment of the 'manifest image' understanding of living beings. One must acknowledge that from another perspective, which is no less legitimate than that of the 'manifest image', life-form/species membership is not absolute, but vague. Thus, this shows that the manifest image of living beings is not isolated from our scientific understanding of them. Rather, it can be adjusted or re-interpreted in response to new scientific knowledge.

Furthermore, this simple reconciliation has been challenged by recent developments in the life sciences. These are known under the label of *evolutionary developmental biology* or *evo-devo*. Discoveries in evo-devo highlight how evolution is enabled by developmental processes, i.e. by processes on the time-scale of individual lifespans. For example, there are multiple levels or kinds of inheritance. For instance, there is 'cultural inheritance' not only among humans, but also among other animals. Animals learn behaviours from their parents, such as tool use among certain monkeys or responses to the particular predators in an environment. Furthermore, there is microbial inheritance, in which microbiome from the digestive system of the parents is passed on to the offspring, enabling them to consume food which they otherwise could not digest. Woodrat mothers in the Mojave desert, for example, feed their newborn some of their own faeces in order to transfer those gut bacteria to them that let them digest the plants in their environment (Andrews et al. 2024).

Very plausibly, traits enabling non-genetic inheritance have the function of passing on adaptations to particular environments while also allowing lineages to adapt quickly to new environments. They have been selected for this. They are widespread within the species in question; they contribute to their survival and they are evidently 'fine-tuned' to enable such passing-on and adaptation. The traits are complex in a way such that their complexity cannot be adequately explained without in some fashion appealing to that effect. Thus, any attempt to interpret, say, desert-woodrat lives, must ascribe this purpose to those traits. Such relatively fast-paced modes of evolution, however, also drive *genetic* evolution. A lineage first becomes adapted to a new environment or niche through cultural and microbial evolution. Then, it acquires genetic adaptations for that environment. These genetic adaptations add up and eventually lead to speciation: the lineage has become a separate species or life-form.

Furthermore, in developing, organisms respond to various factors in order to maintain and increase viability (see e.g. Walsh 2022). This involves 'feedback loops' and reinforcement. Consider, for example, the goat Hopper, a real example discussed by Moosavi (2019). Hopper is born without forelimbs. Growing up, his behaviour and body develop in a way that allows him to move around efficiently, in a way resembling kangaroos. Of course, such capacities depend *inter alia* on genetic information. Nonetheless, genetic information is also among the factors responded to in this way. This is most obvious with novel, potentially harmful mutations. However, all adaptive genetic material originates with random mutations. Hence, the same kind of processes have a role in genetic evolution. It is beneficial to the survival of a lineage and to the fitness of an individual if the lineage can evolve. Thus, these contributions should not be seen as accidental, but as functional. Hence, we cannot understand even momentary 'stills' of evolutionary history without considering evolution. The 'manifest image' of life cannot be isolated from its scientific explanation in evolutionary theory.

From a rather different direction, this conclusion can be supported by considering 'manifest image' judgements about reproduction and cancer. Thompson argues that the judgement that the vital process of reproduction is taking place at a certain time and place (e.g. where this cell is located right now), cannot be reduced to physical and chemical judgements about that spacetime region and its immediate surroundings. This is because physico-chemically identical processes may constitute reproduction for a single-cell organism but growth for multicellular organisms like human beings. In fact, Thompson points out, human cells can grow in a petri dish outside of a human body. It is

possible, if unlikely, that somewhere on earth conditions like in such a petri dish are naturally occurring. In such an environment, colonies of cells chemico-physically indistinguishable from human ones could maintain themselves indefinitely. For these, cell division would indeed constitute reproduction rather than growth. However, the process is chemically and physically indistinguishable from what is happening in human bodies. Thus, so Thompson, judgements about vital processes like reproduction cannot simply be reduced to judgements about physical and chemical processes.

Thompson is right to insist that the hypothetical colonies of cells would not exhibit human life. However, he is too quick to insist that dividing cells within human bodies are not reproducing. It has been a central insight of cell biology that cells in multicellular organisms are self-maintaining units which resemble unicellular organisms in many ways. Furthermore, evolutionary biologists have shown that multicellular organisms evolved from colonies of unicellular organisms. Thus, it is appropriate to describe cell division within a multicellular organism as a form of reproduction.

These insights do not belong merely to the 'scientific image' while being irrelevant to the manifest image. Rather, they affect many people's everyday understanding of life. This can be seen from the everyday understanding of cancer. An important part of our 'manifest image' understanding of life, including our own life, is an understanding of its vulnerability and the many ways it can go wrong. This includes understanding of diseases like cancer. Pre-cell-biological explanations of cancer understood it solely through the category of growth. Cancer was excessive growth caused by some (humoural) imbalance. Through cell biology, we can also understand it as the cancerous cells reproducing. They are doing what cells are always 'aiming' to do. It's just that some regulations have broken down, and now they are not cooperating with the rest of the body anymore. This understanding is reflected, for example, in the way in which cancer patients may think of their tumour as an 'enemy'. Such understandings clearly belong to the manifest image. However, they are only possible if the cell-biological and evolutionary insights have been integrated into the manifest image.

Thus, it is possible for biological innovations to be integrated into the manifest image of life. This leaves Thompson (and Foot) with a dilemma. Either, they reject such integrations, arguing that ethical judgements are based in an 'unreconstructed' understanding of life, or they accept that the relevant 'manifest image' understanding incorporates scientific insights. On the first horn, however,

they are vulnerable to the challenge from science. Compared to a 'manifest image' understanding incorporating scientific insights, the unreconstructed understanding is superstitious or pseudoscientific. In fact, one may suspect that rather than being "perennial" (Thompson 2008 passim, in allusion to Sellars 1962), this 'unreconstructed' understanding of life incorporates *outdated* scientific perspectives, e.g. from Hellenistic empirical biology.

If the 'unreconstructed' logic of vital and natural-historical judgements has been shaped by some --- historical --- forms of empirical sciences, it may also hypostatise certain environments and ways of being human. The historical sciences in question will have incorporated observations of human beings. Thus, their ways of being human would be removed from ethical critique. Plausible examples of this concern reproduction. Neo-Aristotelian naturalism may make it otiose to ask whether one has at least *pro tanto* reason to have children. However, humans may intelligibly ask if there is any value at all for them in procreation.

This worry persists on the second horn. After all, even if the 'manifest image' is being revised in response to natural science, it may still be shaped by these past observations. Furthermore, this option makes neo-Aristotelian naturalism straightforwardly susceptible to the threat of 'biologism'. If the 'manifest image' understanding of life can be shaped by scientific insights, it may also come to incorporate evolutionary understandings of vital function. Thus, it may end up hypostatising either the 'struggle for existence', or some restrictive conception of human nature.

Thus, the distinction between the manifest and scientific image is not enough to show the scientific respectability of the kind of natural-historical judgements purportedly underlying ethics. In order to do so, the neo-Aristotelian must provide another argument to show that scientific biology implicates vital judgements which are not reducible to evolutionary biology.

3 Arguments for natural-historical judgements

In the last section, I have argued that noting that 'natural-historical judgements' belong to the 'manifest image' alone is not enough to respond to the challenge from science. In this section, I critique further arguments that natural-historical judgements form part of scientific discourse.

3.1 Arguments from linguistic indispensability

Thompson regularly refers to the particular grammatical constructions used to describe living beings which, according to him, express vital judgements. Furthermore, Foot (2001) as well as Hacker-Wright (2021) occasionally use the phrase "Grammar of Goodness" to describe their view. These remarks may indicate a response to the scientific challenge along the following lines. (However, they likely also relate to the cognitive/conceptual point of the next section.) Certain linguistic constructions are indispensable for talking about living nature. These may include, for example, generics like 'the domestic dog has four legs' or 'wolves hunt in packs'. Even if such statements are not used explicitly, they may be implied by sentences like 'this is a wolf/dog'. Such constructions, the argument goes, express natural-historical judgements. Thus, natural-historical judgements are indispensable to scientific biology. Since natural-historical judgements provide the basis for ethical discourse, if ethical discourse is pseudo-scientific, so is biology.

We can grant for the sake of argument that scientific biology must rely on generic constructions. However, it does not follow that these constructions must be interpreted as expressing natural-historical judgements. Firstly, some instances of the construction can*not* be interpreted this way. Consider 'the tsetse fly carries the sleeping sickness'. To all appearances, this is an instance of the same linguistic construction as 'the domestic dog has four legs.' However, it cannot be a description of the Tsetse Fly's natural history. The parasite causing sleeping sickness also harms its tsetse-fly vectors. Actually carrying it is a defect in them. Perhaps the construction expresses that the Tsetse Fly is *capable* of carrying sleeping sickness. In other words, it has a natural history which makes it possible for it to carry sleeping sickness. However, dogs presumably have a natural history which makes it possible for them to have the flu. Yet, as a generic, 'the domestic dog has the flu' is not true. Plausibly, any account of 'the tsetse fly carries the sleeping sickness' must appeal to the particular interest humans have in carriers of the sleeping sickness. However, if such an appeal is possible with this sentence, an appeal to, say, human interests in explanation may be relevant to 'the domestic dog has four legs.' This generic may be true because four-legged dogs are good *specimens*, i.e. useful to humans for serving as central cases in describing the diversity of individual dogs.

Furthermore, examples from various languages show that not every idiosyncratic grammatical structure is best interpreted as expressing a distinct 'form of judgement', even if it seemingly

participates in a distinctive pattern of inference. For instance, the Tiwi language spoken on two islands off Australia's north coast, conjugates verbs according to whether an action is performed while walking or remaining in one place (see Osborne 1974). Judgements about activities performed while walking thus have a distinctive grammatical form. Furthermore, they feature in certain distinctive inferences. For example, if a Tiwi verb is conjugated to indicate that the action was *not* performed while walking around, this licenses the inference that the agent was in the same place finishing and starting the action. Nonetheless, these constructions pick out the property of an action 'being performed while (not) walking around', rather than expressing some distinctive form of judgement. (Of course, this property might only be graspable through *vital* judgements; however it does not correspond to its *own* particular form of judgement.)

Similarly, many languages have *grammatical gender*. They mark nouns, pronouns and adjectives as 'masculine', 'feminine' or (in some languages) 'neuter'. Depending on the language, the assignment of grammatical gender with humans and sometimes non-human animals is based on features of the referent. With nouns denoting other referents, it is often arbitrary. Women and females are typically denoted by feminine nouns, males and men by masculine ones. Despite the arbitrariness, grammatical gender is connected to extralinguistic features by some strict rules. Non-compound nouns denoting only females (like 'woman', 'girl', 'queen' and 'actress' or 'cow', 'heifer' and 'sow') are *never* masculine; those that denote only males (like 'bull', 'steer' and 'tomcat' or 'man', 'king', 'father' and 'son') are never feminine[7]. Feminine-gendered pronouns are used anaphorically for

[7] This applies to German, and, to the best of my knowledge, the other gendered languages of Europe. There is a small set of apparent exceptions in German. Firstly, there is a set of slurs which are feminine-gendered but denote queer men. These are likely meant to denigrate men by referring to them as women. Secondly, there is "*der Vamp*" (m.) 'seductress, femme fatale'. However, "*der Vamp*" is an obvious English loanword (even if it is old-fashioned in its homelands these days); it is pronounced with an English-like phonology, i.e. /væmp ~ vɛmp/ rather than */vamp/. So, it may not follow ordinary semantic rules for gender assignment. It is also an 'anglified' clipping of "*der Vampir*" (m.) 'vampire'. So plausibly it inherits its gender from "*der Vampir*". It is further a conventionalised metaphor which describes seductresses as vampires. So the 'literal' meaning would be epicene. (Vampires can be male or female.) It is only the conventionalised metaphorical meaning that is restricted to women. Thirdly, there is "*der Sukkubus*" (m.) 'succubus, demoness'. However, succubuses used to be gender-less demons who merely appear in female forms. King James' *Daemonologie*, e.g., used the older form 'succuba, -ae', but taught that succubuses and incubi were the same creature taking different shapes. The grammatically feminine form "*die Sukkubine*" (f.) isn't found in dictionaries, but it has some Google hits. Thus, speakers feel some tension in the gender of "*Sukkubus*". In any case, there seems to be something uncanny about the few women denoted by masculine nouns.

Some feminine nouns are styles for offices held by men: "*die Heiligkeit*" (f.) 'holiness' and "*die Eminenz*" (f.) 'eminence'. However, these nouns are primarily abstract rather than denoting people. "*Die Obrigkeit*" (f) 'authority, *superioritas*' denotes a person whom a feudal subject owed obedience. Typically, that would have been a man. However, it could also be an institution, or an abbess. Morphologically, "*Obrigkeit*" is an abstract noun. German Wikipedia ("Genus", accessed 13/12/2024) has "*die Memme*" (f.) 'coward, yellowbelly' as a feminine noun only denoting males. It is similar to the recent loan "*die Pussy"* (f.). My native-speaker intuition agrees with leading dictionaries that both can also denote females. If they only denotes males, they seem to belong to the same category as the homophobic slurs (cf. 'sissy').

grammatically feminine antecedent nouns and deictically for female referents. Masculine-gendered pronouns are used with masculine antecedents and male referents. In languages that decline predicative adjectives, such as Slavic and Romance, women use the feminine form in a sentence like 'I am not amused'. Men use the masculine form.

In analogy with generics and natural-historical judgements, it may seem that judgements about gendered individuals constitute their own form of 'gendered' judgement. Such judgements in turn may imply 'gender-judgements' expressible as *this is a s/he*, or *this [goat] is a s/he-[goat]* in much the same way as vital judgements imply natural-historical judgements. Like natural-historical and vital judgments, such judgements give rise to certain inferences. For example, it is an assumption 'baked into' most systems of grammatical gender that, if something is male, then it is not female, and *vice versa*. If I use a masculine pronoun to refer deictically to some individual, I cannot also use a feminine pronoun to refer to the same individual. By the rules of grammar, the reference would have to be to a different individual. Similarly, if a referent is animate, then it is either male or female. Most languages with gender lack non-gendered animate singular pronouns. Furthermore, the grammatical system of gender gives rise to the inference that if an individual has one gender at one point in time, it has the same gender at any other time. In a past-tense sentence like '*he* sat on the chair', the gender of the pronoun relates both to the referent's extralinguistic gender/sex at the time of the described event, and to his gender/sex at the time of utterance. There is no grammatical way to express that the referent has changed gender in the meantime. In addition, gender systems typically interact with vital judgements, specifically judgements about reproduction. With these gender systems, a judgement like ***she*** *is his parent* entails *she is his* ***mother***. (Let us assume that in the language in question, the concepts denoted by ordinary translations of 'parent' and 'mother' apply equally to humans and other animals.) *Mother* is simply the feminine-gendered analogue of *parent*. Yet, it also

Etymologically, *"Memme"* is a dialectal word for a woman's breast. It came to denote a child crying for its mother's bosom, and then a coward. If the anatomical meaning is still primary, this also explains the surprising gender.

Other than that, nouns denoting both males and females (like 'person', 'human', 'child' or 'goat') may be masculine, feminine and neuter. German diminutives in "*-chen*" and "*-lein*" are always neuter, hence: "*das Mädchen*" (n.) 'girl', "*das Männlein*" (n.) 'little man' etc. A few non-diminutive nouns denoting only women are neuter too, e.g. "*das Weib*" (n.) 'woman' (obsolete, now only derogatory) and "*das Mensch* [sic]" (n.) 'woman' (now regional and typically derogatory). (There seem to be no neuter non-diminutive, non-compound nouns denoting only males. This may well reflect a sexist assumption about females being less agentive.) There are some compound nouns which have masculine gender and female referents or *vice versa*, e.g. "*der Weibsmensch*" (m.) 'lit. the she-human' or "*die Mannsperson*" (f.) 'lit. the man-person' (both old-fashioned). However, in those cases the head is epicene. The compound's gender is determined by the head. The latter also applies to a handful of figurative compounds: "*der Blaustrumpf*" (m.) 'bluestocking', "*der Backfisch*" (m.) 'deep-fried fish; female teenager', "*der Wildfang*" (m.) 'wild-caught animal; tomboy', "*die Sexmaschine*" (f.) '(of men) sex machine'. All these later examples are straightforwardly compatible with the generalisation that non-compound nouns denoting only females are never masculine and *vice versa*.

has more particular implications. Given the assumptions built into vital judgements, *she is his mother* in turn entails *she gave birth to him or laid the egg from which he hatched.*

Such linguistic data may be taken to support an account of human gender according to which it is constructed through our practices, including linguistic practices. However, whatever the merits of such an account as regards human beings, it is implausible regarding other animal referents. While admitting of vagueness, animal sex is a biological property if any organism-level property is. This property is picked out by grammatical gender in the case of non-human animal referents. Tiwi features a five-way gender contrast: [HUMAN, FEMININE]; [HUMAN, MASCULINE]; [NON-HUMAN, FEMININE]; [NON-HUMAN, MASCULINE] and [INANIMATE] (see Osborne 1974; a similar system is found in the unrelated Australian language of Anindilyakwa, see Leeding 1989). Thus, despite being deeply embedded in the grammar, the contrast between [NON-HUMAN, FEMININE] and [NON-HUMAN, MASCULINE] simply picks out a property that can be reduced to judgements which can be expressed without relying on that particular linguistic construction.

Given this, some empirical discoveries can lead to the rejection of inferences licensed by the grammatical form. For example, some animal species, such as slugs, do not have sexes. All typical specimens produce both spermatozoa and ova. In some species, individuals may change sex during their lifetime. Typical male clownfish, e.g., will develop into females if there are no other females around. Rifkin and Garson (2023) argue that many animals, including humans, have more than two biological sexes. Biological sex, they argue, is determined by which kind of gamete (spermatozoon and/or ovum) an individual is *selected for* producing. Some individuals are not selected for reproducing directly at all, thus they are not selected for producing spermatozoa *or* ova. Similarly, reproductive technologies show 'biological' mother- and fatherhood to be complex. There may be three different individuals supplying the ovum, the maternal DNA and the womb for one embryo.

In response to such discoveries, speakers can consciously attempt to adjust their grammar. They may introduce new constructions which are at first ungrammatical, but which they hope will through use become part of their language's grammar. They may also start using gendered language in a self-consciously fictional or 'as if' way. For example, a community of Tiwi-speaking zoologists may agree to refer to all slugs using [NON-HUMAN, MASCULINE] forms. In doing so, they cancel the

implication that the slugs are all male (and not female). Similar strategies have of course been pursued with regard to gender and *human* referents both in English and in many other --- more thoroughly gendered --- European languages. Such changes may or may not also be a response to empirical discoveries about a property picked out by gendered linguistic constructions. Languages in other parts of the world, too, likely long had patterns of using gendered language that are different from the 'traditional' European one. They are similarly subject to change. The use of expressions like *mother* and *father* has also changed in relation to institutions of step-, foster and adoptive parenthood.

Hence, the generic statements used in science may be similarly explainable, as picking out certain complex properties which can also be picked out without special forms of judgement. Insofar as generic judgements license special inferences not licensed by other forms of judgement, it may either be rational to change the grammar of our language, or to use the constructions that enable such inference in an 'as if' way, so that the implications in question are cancelled. Thus, they do not indicate that biology depends on any particular form of judgement. Hence, firstly, any form of ethical discourse based on generic statements about living beings that are *not* explainable as denoting some property may well be superstitious. Secondly, such an ethical discourse would hypostatise the grammatical forms it depends on. As the foregoing shows, speakers can deliberately shape the way their language's grammar develops. Thus, such a grammar is subject to ethical critique. However, a form of ethical discourse based in such grammatical forms would remove them from such criticism.

3.2 Arguments from cognitive indispensability

In addition to linguistic arguments, there are also *cognitive* arguments for the indispensability of vital and natural-historical judgements. These are arguments that, independently of how their results are expressed linguistically, life-sciences would not be possible without relying on natural-historical judgements. Thompson firstly argues that the biology could not recognise its domain without relying on such judgements, i.e. it could not distinguish living from lifeless nature. Secondly, he argues that empirical biology needs to rely on vital judgements, and thus indirectly on natural-historical judgements, for its data. The data of empirical biology are observational judgements about living individuals. Such judgements, so Thompson, however, are vital judgements. I argue that we can grant Thompson's empirical descriptions of how biology is practiced. It does not follow that mature biology depends on vital judgements as conceived by Thompson.

Thompson develops the first argument by drawing on a list of criteria of living things drawn from a biology textbook. He notes that each criterion depends on vital judgements. For example, living beings are said to 'reproduce' in contrast to lifeless nature. However, so Thompson, the relevant sense of 'reproduce' is a vital description. The same is the case with 'nourish', etc. Yet, Thompson does not consider sophisticated, formal definitions of 'life' which have been developed, e.g., within the Astrobiology and Artificial Life research traditions (e.g. Ruiz-Mirazo et al. 2004, Mix 2020)

Thompson might respond that biology cannot be guided by such definitions since most biologists are not familiar with them. However, the development of those definitions can itself be seen as progress in biology. Earlier observations in biology did indeed depend on vital judgements as described by Thompson. However, such judgements merely represent one theory in biology. This theory is being revised in light of empirical evidence, leading to the newer theories expressed in these definitions. Thus, mature biology does not depend on vital judgements as described by Thompson. Analogous accounts can be given for judgements about individual living beings.

The foregoing resembles holistic accounts of the epistemology of science, i.e. about how scientific theories are tested and supported. Such views have been presented e.g. by Quine (1951), Popper (1959) and Lakatos (1968). It can be supplemented by different views about the *semantics* of scientific theories, i.e. the question how they relate to their subject matter. It is *compatible* with a nominalist, Quinean scepticism concerning the relation between any parts of the theory and a world as it is 'in itself', independently of the theory. However, it can also be combined with substantial semantical theories of the relation between scientific theories and their subject matter. It is compatible with the equivalents of Cornell-style and 'Canberra plan' naturalism.

According to the 'Cornell-style' reading, vital judgements approximately have a particular causal connection to a natural kind. They approximately *track* that natural kind, i.e. most of their instantiations are caused by the natural kind. A natural kind in the relevant sense may be a "homeostatic property cluster" (Boyd 1999). That is, it may be an assemblage of interrelated and reinforcing properties. The natural kinds in question constitute being alive, being an individual organism, being a life-form/species etc. In virtue of this tracking relation, these judgements picked

out that property. Through advances in biology, the revised concepts of life etc. allow us to describe that property more accurately and track it even more precisely.

According to a functionalist or 'Canberra plan' reading, the 'logic' of vital and natural-historical judgements, i.e. the inferential relations between them, can be expressed as platitudes. These platitudes constitute an implicit *theory* of living nature. The best fit for this theory is provided by certain properties or natural kinds. These are the properties picked out through vital and natural-historical judgements. However, that fit is not *perfect*. Thus, some aspects of the theory are false. Through scientific progress, we develop theories which the same properties fit even better.

Thus, the critic of neo-Aristotelian naturalism may grant that vital judgements are indispensable for scientific biology in a historical sense. Scientific biology could never have gotten off the ground without relying on them. The same may be the case in individual biographies. Perhaps a learner must first grasp a less accurate theory as embodied in everyday judgments about living nature before they can understand scientific theories about it. However, this does not mean that scientific biology cannot lead us to revise vital judgements, and even to revise their structure in fundamental ways. Thus, neo-Aristotelian naturalism still faces the dilemma of Section 1. It can insist that the notion of life-form relevant to ethical judgements is not affected by scientific insights. In that case, it is open to the challenge from science: ethical judgements rest on superstition. It is also open to the worry that it hypostatises ways of being human that are preserved in the logic of vital judgements from past observations. Alternatively, they can admit that the relevant notion integrates new scientific knowledge. If so, it appears that it must integrate evolutionary notions of function. Thus, neo-Aristotelianism would provide no response to the threat of biologism.

3.3 Internal scientific arguments

Even if there is no 'manifest image' of living beings that is unaffected by scientific biology, neo-Aristotelian naturalism may succeed. This is because mature biological theories may feature a notion that can play the role of life-forms in vindicating ethical discourse. Arguments to this effect are presented by Walsh (2022) and Moosavi (2019).

Walsh argues against the 'Modern Synthesis' of population genetics that the organism is a relevant unit of explanation for biology. This is, inter alia, because it is the activities of organisms which

enable the inheritance of genes, and which produce adaptations by responding to genetic mutation. In order to understand organisms, however, we need the analogue of Aristotle's notion of *bios*. Walsh (2022: 289) translates *bios* as "way of life" or "characteristic way of life" and connects it to the organism's "goal-directed capacity to direct the matter at its disposal toward the attainment of a well-functioning organism that is typical of its kind". Thus, the human *bios* may be identified with the human life-form according to Thompson. Ethical discourse is not superstitious since it is continuous with discourse about *bioi* that has an integral role in scientific biology.

However, there are problems with Walsh's approach. Firstly, his arguments do not show that scientific biology relies on the notion of a *species*' characteristic way of life. Even if we can only make sense of evolution by considering the behaviour of individuals and their interactions with their environment, it may well be that scientific biology only needs a notion of a pattern of behaviours and interactions in the *individual* case. They may then generalise from the individual cases to species by relying, say, on statistical normality or evolutionary considerations. A species' *bios* may simply be the *bios* of a species-typical individual, i.e. of an individual realising its traits' selected-for functions.

Secondly, it cannot avoid the threat of biologism. Being a theoretical concept, Walsh's *bioi* are completely determined by their role in theory. The theory, in turn, is determined by its role in explaining or predicting observations. Thus, the content of the human *bios* is determined by its role in making sense of human behaviour as it appears in empirical, biological observations of human beings. Thus, just like evolutionary naturalism, an account of ethical discourse based on judgements about *bios* would hypostatise a certain way of being human.

In contrast, Moosavi argues that scientific biology must include judgements about the *welfare* or *thriving* of living beings that are not reducible to other judgements in scientific biology. Ethical discourse could thus be construed as being about human welfare. She considers the aforementioned case of Hopper, a goat born without forelimbs. Hopper grew up with a unique morphology and a unique pattern of behaviour, which allowed him to survive and move around efficiently. According to Moosavi, evolutionary accounts (as well as non-individualist neo-Aristotelian accounts) do not allow us to ascribe any functions to the goat's unique morphological and behavioural features. After all, they did not evolve. This, so Moosavi, is untenable. Hopper's features are evidently functional.

If function ascriptions are not based on evolutionary considerations, there must be some other way of distinguishing them from purely accidental benefits. For example, I may be lost in an avalanche, and my heartbeat alerts rescue dogs to my position. Still, making a noise to indicate my location is not a vital function of my heart. In order to differentiate vital functions from accidental benefits, Moosavi appeals to accounts that connect biological function to self-maintenance (e.g. Moreno and Mossio 2015, McLaughlin 2002). According to such accounts, a trait has a function insofar as it has some effect through which it contributes to the self-maintenance of the organism in the constant exchange with its environment. Avalanches and rescue dogs are not an ongoing feature of my environment. Thus, the exchange with the environment in which my heart alerts the rescue dogs is not a moment of my mode of self-maintenance. In contrast, my heart's pumping blood is such a moment.

However, Moosavi argues, such accounts of function cannot completely reduce vital function to empirical biology. Instead, they must rely on an irreducible notion of welfare or the good. Otherwise, they face the 'overgeneralisation problem'. Consider the following cases. An obese person avoids exercise because their weight makes it arduous. Thus, they maintain their (high) weight. Someone's high anxiety causes them to perceive many potential dangers in their surroundings, thereby keeping their anxiety levels up. The obesity/anxiety are part of self-maintaining systems. Yet, they are not pre-theoretically functional. According to McLaughlin's (2002) solution, which Moosavi endorses, the anxiety (or obesity) is not functional because it does not contribute to its bearer's *welfare*. Thus, judgements about function imply judgements about welfare. Ethical discourse is continuous with such judgements, and thus not superstitious.

However, Moosavi's argument is unconvincing. Firstly, scientists may ascribe functions to Hopper's traits on the basis of evolutionary considerations. Goats (like most complex organisms) likely have an evolved (and characteristic) meta-capacity to develop in ways enabling them to move around comfortably given some fixed conditions. This capacity will involve some feedback mechanisms and play a role in the development of typical as well as atypical body-plans. Hopper's unique morphological and behavioural patterns will have (derived) etiological functions insofar as they are the product of such a meta-capacity.

Furthermore, Mossio and Moreno (2015) have responded to the overgeneralisation worry. Firstly, they suggest that the traits in question (obesity, anxiety) are not integrated tightly into the organism. Thus, it is only the obesity/low-exercise complex that constitutes its own small self-maintaining system rather than belonging to the organism as a whole. Secondly, Moreno and Mossio have stressed the importance of higher-order regulatory subsystems. Obesity is not regulated for, but rather regulated *against*.

Finally, the kind of judgments about function Moosavi describes are not integral to scientific biology. While it is 'counter-intuitive' to say that Hopper's traits are not functional, nothing else in scientific biology depends on judging that they are. Thus, it may well turn out that scientific biology requires us to say that Hopper's traits are non-functional. Scientific biology has compelled us to accept more surprising conclusions before. Human beings share a common ancestor with snails! Thus, Moosavi's arguments do not show that ethical discourse is scientifically respectable. Rather, the kind of discourse about function that she describes may be dismissed as no less superstitious than astrology --- or ethics.

3.4 First-personal arguments

Thompson (2004) presents these three theses (along with two others):

> A mature human being is typically in possession of a non-empirical singular representation of one individual organism.
>
> Individual human beings are sometimes in possession of non-observational knowledge of contingent facts about one individual organism.
>
> Human beings are characteristically in possession of some general substantive knowledge of the human life form which is not founded empirically on observation of members of their kind, and thus not 'biological'.

This may be read to indicate some particular 'first-personal' source of knowledge concerning the human life-form, i.e. a source of knowledge about myself, which is accessible only to me, and from which I can generalise concerning the human life-form. (Whether this is Thompson's intention

depends on how the last thesis relates to the earlier thesis that the "[t]he concept human, as we human beings have it, is an a priori concept attaching to a particular life form".) Such first-personal access may, for example, consist in introspection, or phenomenological reflection *à la* Merleau-Ponty (2012 [1945]) as suggested by Drummond (2002, 2013, 2015). Through such means, I may know that there are at least some true vital judgments, i.e. those concerning me, and at least some true natural-historical judgments, i.e. those concerning my life-form (whatever life-form that may be). Hence, so that argument might go, I can reject any scientific worldview which does not allow for the truth of such judgements. Thus, the successful scientific worldview must allow for at least some true natural-historical judgments. If it allows for some true vital and natural-historical judgments, it can allow for others at no additional ontological 'cost'.

However, firstly, it is not clear that such first-personal awareness supports natural-historical judgments. (In)famously, Descartes' (1998 [1641]) introspection only reveals a thinking substance, which is not necessarily alive. Even for Merleau-Ponty (2012 [1945]), phenomenological reflection only reveals my own living body, and an "anonymous body-schema" which is mediated by my particular society. It does not reveal a universally human 'life-form' or species.

Secondly, as such differences in result underline, first-personal methods are not infallible or incorrigible. Through empirical evidence, I may come to regard first-personal (introspective of phenomenological) seemings as illusions. Such seemings may even change if I internalise such evidence. Thus, even if I seem to have first-personal awareness of my life-form, I may come to see this as an illusion due to the development of science.

Finally, this argument would give rise to a 'problem of other lives', analogous to the problem of other minds. Biological science shows that the observable behaviour of all living beings can be explained without appealing to vital and natural-historical judgements. Yet, according to the neo-Aristotelian approach, such judgements are also not entailed by other empirical judgements. Thus, my observations of other (seemingly) living beings provide no evidence for vital judgements concerning them. Thus, even if I do not judge my seeming awareness of my own life-form to be illusory, for all that the first-personal arguments could ever show, my inclination to judge that other beings bear life-forms is misleading. Thus, it turns out that, for all I am justified in believing, I am the only living being, the only actual bearer of my life-form.

This is not only highly counter-intuitive. The neo-Aristotelian argument that, say, promise-keeping is virtuous depends on the assumption that I share a life-form with my promisees. If they are lifeless automata, and I am the only living being, then I do not need to keep my apparent promises.

One might object that it is an a priori truth that members of a life-form must be descendants of other members of the same life-form. Thus, there must be some other human beings. However, firstly, this is not an a priori truth. It is no truth at all. Hybrids, such as mules, are the offspring of two different life-forms. For all I know first-personally, I am a hybrid. My parents might belong to different life-forms. If so, I would not be bound to them by any obligation. In any case, there is only a small number of people with whom I know my blood-relation. For most people, I rely on biological science to know that we share a common, human ancestor. According to the argument under consideration, however, biological science treats living beings as if they were lifeless automata. Hence, it cannot provide me with evidence that others share my life-form.

3.5 Transcendental arguments proper

Finally, neo-Aristotelians might suggest that natural science must admit vital and natural-historical judgments in order to explain its own possibility. Such an argument may be called a *transcendental argument*. According to it, science is itself a normative or norm-guided practice (cf. MacIntyre 1988 on intellectual virtues; McDowell 1998). One can be a good or a bad scientist. Scientific theories are supported by reasons to believe them. Thus, natural science is inconsistent if its worldview leaves no room for evaluative facts or facts about reasons. Its own existence falsifies its worldview, and an agent could not consistently subscribe to the worldview while being aware of doing so. The neo-Aristotelian naturalist may continue that the best (or only) account that natural science can give of evaluative and normative facts involves neo-Aristotelian naturalism. Science is a rational human activity. As such, it can be excellent or defective. The normativity attaching to it is a kind of human rational normativity, which is a kind of vital normativity.

However, firstly, simply stating that natural science must give an account of itself as a normative activity is not enough to actually give such an account. As we have seen in the preceding subsections, a good scientist, i.e. a scientist following the norms of science, would not accept natural-historical or vital judgements into her theories. Thus, the transcendental argument may locate a contradiction in

the scientific worldview. However, it does not show that neo-Aristotelianism is a way to resolve that contradiction. If neo-Aristotelianism is incompatible with the scientific worldview, as I have argued above, then conjoining the two does not resolve the contradiction. It merely replaces it with another.

Furthermore, the transcendental argument begs the question against biological naturalism. If biological naturalism can successfully explain the nature of normativity, then the scientific worldview is not devoid of norms. One might argue that the threat of 'biologism' and hypostatisation makes a biological account of science's normativity impossible, just as it seems to make an analogous account of ethical normativity impossible.

However, it is not clear that such an account of science's normativity must resist hypostatisation in the same way as an account of ethical normativity. The normativity of science is primarily about the normativity of belief or similar cognitive states. However, for example, the question 'this is a descriptive belief which a species-typical *H. sapiens* would hold, but still: why believe it?' may really turn out to be pointless. Epistemic warrant or truth may plausibly be identified with proper functioning (cf. Plantinga 1993; Millikan 2004). Questions like 'this is true, but why believe it?' and 'this is the cognitive state I'd be most justified in holding, but still, why hold it?,' however, are pointless.

There are, of course, conflicts between our pre-theoretical judgements about truth and justification and such identifications. Some false beliefs may be selected for, e.g. beliefs arising from misremembering giving birth as less arduous and painful than it was. Similarly, some epistemically unjustified beliefs may be selected for. For example, it may have contributed to our ancestors' fitness to overestimate how much their success in their projects was up to them compared to what their evidence supported. However, these worries can all be explained as worries about the particular extension. If their empirical assumptions are correct, they show that the extensions of *true/justified belief* are too far from those of, say, *belief actualising its selected-for effects* or *belief produced by a well-functioning process*. Thus, it can't be that the latter denote the same kind or property as the former. However, if there should be a concept derived from biology that matches the extension closely enough, then there would not be any additional worries that such an identification would distort the practice of science. Science, and epistemic pursuits more generally, are not exercises of freedom in the same

sense as ethical discourse. In science, reason may "consciously respon[d] to a bidding from the outside".

4. Conclusion

In this chapter, I have introduced neo-Aristotelian naturalism and argued that neo-Aristotelian naturalism fails to meet the challenge from science introduced in Chapter Two. In the following chapter, I will propose a form of biological naturalism which can meet them.

FOUR: From the organisational account of virtue…

In this chapter, I will outline the *organisational account of virtue*. This is an account of virtue based on the organisational account of function, i.e. the account of function based on self-maintenance and self-regulation which I introduced briefly in Chapter One. In the following chapters, I will develop the organisational account into *bio-constructivism*: a more complete account of ethical inquiry and ethical truth. In doing so, I will argue that the bio-constructivism/the organisational account of virtue can avoid 'hypostatisation' as described in the preceding chapter (Chapter Two; see also Chapter Four) and that it can explain the connection between ethical judgement and action (Chapter Five).

In this chapter, I will first set out the organisational account of biological function. Then, I will describe how it applies to the systems which determine behaviour in humans. I will introduce the notion of different levels of self-maintenance, and of a *general regulatory system*. This is a system which advances self-maintenance on *all* levels. *Virtue-inquiry,* i.e. the component of ethical inquiry concerned with which considerations to treat as reason, is such a system regarding human behaviour. The subsystem(s) regulated by it constitute an agent's *character*. Its components and features are *character traits*. Character traits have biological functions. A character trait actualising its biological function is a *virtue*.

1. The organisational account of function

The organisational account is based on the intuitive notion that organisms are systems that endure because they maintain themselves by interacting with their environment and changing their material composition. Functions are the contributions that traits make to such self-maintenance. The idea that self-maintenance is central to living systems has been influentially defended by Maturana and Varela (1980). It has been applied to the problem of biological functions among others by Schlosser (1998), Collier (2000), Bickhard (2000, 2004; Christensen and Bickhard 2002). However, I draw primarily draw on the recent formulation by a group of biologists and philosophers centred around the University of the Basque Country: Mossio et al. (2009), Saborido et al. (2011), Nunes-Neto et al. (2014), Moreno and Mossio (2015), Saborido and Moreno (2015), Saborido et al. (2016), Mossio and

Bich (2017), and Mossio and Pontarotti (2022). This group has coined the expression "organisational account of function".

Following Saborido et al. (2011), Moreno and Mossio (2015: 73) define *function* as follows: "a trait T has a function if, and only if, it exerts a constraint subject to closure in an organisation O of a given system. This definition implies the fulfilment of three different conditions:

C1. T exerts a constraint that contributes to the maintenance of the organisation O;
C2. T is maintained under some constraints of O;
C3. O realises closure."

A *constraint* is an effect that makes other events more likely, even if it is not strictly necessary for them to occur. For example, a given organelle in a cell may exert a constraint by producing an enzyme which acts as a catalyst for some other reactions. The reactions in question can in principle occur without the catalyst, but on the relevant time-scales they are much less likely. A constraint is subject to *closure* iff the following is the case: It is an element of a set of constraints. Each of the elements of this set is produced under some other constraint which is an element of the set, i.e. it is made more likely by that constraint. Each element of this set also makes some other element of the set more likely. Furthermore, each element of the set must be related to each other element of the set through a series of links such that each element directly or indirectly contributes to the maintenance of each element. This closed set of constraints and their interrelations is called an *organisation*. Thus, for single-call algae like *chlorophyta*, the presence of sunlight is a constraint necessary for photosynthesis to occur, but it is not part of its closed organisation since it is not produced under some constraint that is part of that organisation. Similarly, the noise being made by a mammalian heart might act as an external constraint on some other organism, such as a predator listening for it, but it is not part of the organisation if no other constraint depends on it.

Under this account, we can see that a heart is functional because it exerts a constraint --- constant blood-flow --- under which many other functional traits of an organism operate. In turn, it is maintained under some other constraints produced by the organism. For example, it would become necrotic if not supplied with oxygen through blood. Even though it is not explicit in the definition,

we might add that *the function of a trait* consists in the constraint(s) it exerts under which other constraints are exerted. So, the function of the heart is to maintain constant blood-flow.

Sometimes, a failure of function might lead to a breakdown of the containing system, as with the heart. However, sometimes it may only lead to a change of organisation. For example, if the kidneys of a contemporary resident of an industrialised nation are failing, she might receive dialysis. Thus, the organisation would have changed. It now does not feature the constraints exerted by the kidneys, and it does depend on the external constraints exerted regularly by the dialysis machine. Some such corrections might not depend on the environment to the same degree. A given muscle in my leg may be functional because it allows me to move around and get food. If that muscle was failing, I could adopt a different gait. This would be a different organisation. Different muscles would exert different constraints.

One common objection to the organisational account of function relies on cases that have been described by Boorse (1976) and applied to the organisational account by Garson (2017), which we discussed in Chapter Five. They aim to show that the organisational account over-ascribes functions. Imagine Bob. Bob is obese and abstains from exercising because it is made arduous by his body fat. He retains his body fat because he doesn't exercise. His body fat and his non-exercising form a self-maintaining system. Thus, seemingly, the function of his body fat is to prevent him from exercising.

However, in this case, the body fat only has a function as part of the fat/non-exercise system. The constraint it exerts in making it arduous to exercise is not part of the highly complex closed organisation that most of Bob's functional traits belong to. Defenders of the organisational account may elaborate their definition to the effect that O must have some degree of complexity that excludes such cases. Mossio et al. (2009) for example already require that O must be an organisation of a system S that is organisationally differentiated. Different parts of S must exert different constraints that contribute to self-maintenance in different ways. This is intended to exclude very simple self-maintaining systems such as flames. However, a similar condition may also exclude the fat/non-exercise system.

Another important challenge to the organisational account consists in the contention that it cannot explain malfunctions or dysfunction, i.e. cases where a trait has a function despite failing actually to perform it in a given organism. Suppose my kidneys are not filtering toxins out of my blood. Instead, I regularly undergo dialysis. Still, filtering toxins out of my blood is the function of my kidneys. They have a function even though they make no contribution to the self-maintenance of the organism.

However, Moreno and colleagues provide an account of malfunction (Moreno and Mossio 2015, Saborido and Moreno 2015, Saborido et al. 2016). According to Moreno and colleagues, not all traits that satisfy C1-3 can exhibit malfunction. Rather, for a trait to count as (potentially) malfunctioning it must belong to a system that has *regulatory subsystems* which respond to failures of traits to exert constraints "*presupposed*" by other functional traits in the system. A functional trait presupposes a constraint in the relevant sense if it contributes to self-maintenance more effectively if the constraint is present. This is simply a counterfactual relation. *If* the presupposed constraint would not be exerted, *then* the presupposing trait would contribute to self-maintenance less effectively. The effectiveness of self-maintenance can be described as the system's *viability*. The greater the system's viability, the longer is its expected duration of survival in the kinds of environments it is likely to encounter. For example, thrombocytes respond to the failure of blood vessels to contain blood by causing coagulation. Some regulatory subsystems of this kind respond to many different failures. For example, if I or a similar animal have a torn tendon, I may feel pain and respond by consciously and unconsciously adjusting my movements until I feel less pain. The responsible subsystem is part of my mind. It can respond to many different kinds of failures in many different muscles in different appropriate ways.

Now, a trait T is malfunctioning iff the following conditions are fulfilled: (1) T is functional; (2) T fails to exert some constraint to the extent presupposed by other functional traits of the same system; (3) a regulatory subsystem responds to this failure; (4) the response of the regulatory subsystem does not shift the system to an organisation that is as effective as it would be if T did not fail to exert the presupposed constraints.

(1) is necessary to exclude sources of external constraints. The sun is not malfunctioning during a solar eclipse even if it fails to supply plants with the light needed for photosynthesis. (2) captures an intuitive core of the malfunction.

(3) is necessary since otherwise, almost any trait of any system would count as malfunctioning since under some conditions, however unusual, it might contribute better to self-maintenance. The melanin in our skin contributes to self-maintenance by protecting us from UV-radiation, i.e. by exerting the constraint of low rates of UV-radiation within deeper layers of the skin. This constraint is presupposed by other functional traits, e.g. those involved in copying DNA. In fact, these traits can be said to presuppose the total absence of UV-radiation. If there was no such radiation, their contribution to self-maintenance would be optimally efficient. If the presence of UV-radiation in our environment was negligible, the melanin could exert the constraint of UV-radiation being completely absent. However, its failure to do so under normal conditions does not count as a malfunction.

(4) is needed to preserve the intuitive core of *malfunction*. Why should a trait count as malfunctioning if it doesn't make the system self-maintain any less efficiently? If my environment gets colder and I wear thicker clothing to correct for the 'failures' of my thermoregulation, my thermoregulation doesn't malfunction.

If a trait has a function which is capable of malfunction in this way, even if it is not currently malfunctioning, we can say that it has a *regulated function*. If a trait systematically and consistently contributes to self-maintenance but is not subject to regulation and thus not capable of malfunction, it has a *bare function*.

This account can be described the following way. A self-maintaining system can have relations to its environment such that *for it*, the environment *ought to be* a certain way. For this, minimally the system must be such that some of its functional traits presuppose the presence of a certain constraint. However, there is a stronger sense of 'ought for'. Many organisms' traits presuppose the presence of gravity. Yet, there is a sense of 'ought for' where it is inappropriate to say that, for them there ought to be gravity. Thus, we can add the following condition. The organism must respond to the failure of that presupposed constraint. For example, if there is low glucose in their environment, *E. coli* bacteria move randomly until they reach an area with higher glucose. Thus, we can say, in a stronger

sense of 'ought for', that for an *E. coli* bacterium, there ought to be glucose in the environment. The organisational account of malfunction applies this notion reflexively. An organism's trait is malfunctioning if it fails to be as it ought to be for the organism.

2. The functionality of character traits

In this section, I show that character traits as conventionally conceived can be functional, i.e. that they can satisfy the conditions C1-3 of the organisational account of function. In the second subsection, I will focus on C1. First, I focus on C2-3. In doing so, I will develop a suggestion of character traits as features of a high-level subsystem of human organisms. In Section 2, I will say more about what distinguishes character traits from other mental features of organisms under the organisational account of virtue. For now, however, I will rely on the commonsense notion of character traits, and the notion of character traits found in different forms of 'function-based naturalism' described in Chapters Two and Three. Character traits are complex, enduring features of human agents which determine their behaviour. Taken together, an agent's character traits constitute their *character*. Character is the particular shape that *practical reason* takes in an agent. Practical reason is the capacity to treat considerations as reasons for action. An agent's character is a matter of which considerations that agent treats as reasons for action.

2.1 Character traits are parts of self-maintaining systems

In this section, I argue that character traits satisfy conditions C2-3, i.e. that character traits are components of self-maintaining systems. In some sense, this seems obvious. John's shyness, say, does not continue to exist if John ceases to exist. Hence, John's shyness is maintained by the organisation of the human organism known as 'John'. This organisation realises closure.

However, one might still wonder if character traits like shyness are the kinds of 'traits' that can be described as functional. In the following, I will consider some such worries. For example, one might ask whether 'character traits' are traits that are maintained under particular constraints. With a trait like the heart, it's easy to specify some of the particular constraints that are required for self-maintenance, such as a protection from external forces and a supply of oxygen and sugars

through the bloodstream. Other constraints are irrelevant. In contrast, it's not easy to qualify any particular constraints under which character traits might be maintained.

However, such partial ignorance does not mean that we cannot know that character traits are not maintained under some constraints. Some constraints are clearly *ir*relevant to the maintenance of character traits. For example, the presence of certain enzymes in the stomach. They may indirectly impact character traits. However, they do so by impacting the ways other constraints are produced. For example, a certain configuration of enzymes may cause constant low blood sugar and thereby make you more irritable.

For a comparison, we can consider thermoregulation. Scientists may discover that some species of an otherwise ectothermic genus displays some endothermic thermoregulation. That is, they produce some heat within their body through metabolism in order to maintain a certain body temperature, which is necessary for self-maintenance. Even without knowing anything about how the mechanisms bringing about this thermoregulation operate, we can conclude that their producing the body temperature is functional.

One may also object that character traits do not satisfy C2 because they are only emergent properties of a human organism. At most, the objector may suggest, the components of the system which together constitute the character trait are maintained under constraints exerted by the relevant organisation.

'Emergent property' may mean different things. The objector cannot simply mean that character traits depend on or are constituted by other components of the organism. Systems can be described at different levels. When we ascribe a function to the stomach, we are not considering the constant processes going on in the individual muscle cells that constitute it. On an even higher level, we may describe the whole digestive system as having the function of introducing nutrients into the bloodstream.

In fact, we can approximately describe most vertebrate animals as self-maintaining through the interplay of 10 'body systems' (skeletal, muscular, nervous, endocrine, cardiovascular, lymphatic, respiratory, digestive, urinary, reproductive; see e.g. Keener and Sneyd 2009). Each of these systems

can be ascribed some major functions. Such a description not only abstracts away from the operations of each system, but also from the more subtle functional interconnections between aspects of different systems. Nonetheless, it is useful and appears to capture some biological reality. Given that the organisational account can easily accommodate such function ascriptions, it should not be interpreted in ways that exclude them.

Many have thought of the 'soul' or the 'mind' as composed of different parts that interact and produce effects. In the *Republic*, Plato presents a view of the soul as consisting of three parts, sometimes translated as *reason*, *spirit* and *appetite*. In a crucial allegory, Reason is presented as a charioteer, driving the horses Spirit and Appetite. The notion is not confined to Western traditions. Ancient Indian philosophers see character as constituted by a combination of three *guṇas* 'elements' or 'substances': *sattva* 'honesty', *rajas* 'passion' and *tamas* 'darkness' (see Nadkarni 2014). The *Katha Upanishad* (third chapter, composed between the 5th and 1st century BCE), features a chariot analogy resembling Plato's. The mind (*atman*) is the charioteer, while the 'senses' are the horses.

Later, thinkers as diverse as Kant and the psychoanalysts defended similar views (see Harcourt 2015). The same is true of many recent and contemporary cognitive scientists. For example, it applies to modular models of the mind. According to such, the mind consists of many specialised modules (see Philip 2017). It is also suggested by theories like that of *emotion regulation* according to which humans manage their own emotions through various strategies (MacRae and Gross, 2020).

Such *multipartite* models of the mind can be seen as high-level descriptions of a (sub)system within human organisms. This high-level system is instantiated through the interaction of low-level systems. This is similar to how functionalism in the philosophy of mind treats the states ascribed in commonsense mental activity, such as beliefs and desires (see Lewis 1972). Note, however, that my position does not depend on functionalism being correct. Functionalism in the philosophy of mind implies that *all it takes* to have a certain mental state is to be in a state exhibiting the relevant functional profile. I am assuming only that mental states instantiate the structure described in those models, not that this is sufficient for being a mental state. Thus, it may be, say, that mental states must be instantiated in a particular material way or that they may need to have non-physical phenomenal properties. They may also need to be in a certain physical relationship to the external objects they are about.

Differing models of the mind may simultaneously be apt. Consider for instance those proposed by modularist cognitive scientists, who assume a large number of modules, and by moral psychologists in the Platonic tradition, who assume three parts of the soul. They may describe overlapping or identical subsystems at different levels. They may also describe different approximations to that subsystem, highlighting and passing over different aspects of it. I will not rely on a particular multipartite model of the mind, but rely on features that many such models have in common.

Given this, character traits may be components of such a system, features of its components, or features of their configuration. In each case, they can be appropriately described as being maintained by the system, and as exerting constraints. Thus, they are potential candidates for having functions.

This fits with the way character traits are described, for example, by Hursthouse (1999). She suggests that they are patterns in ethical inquiry, perception and emotional responses which support and reinforce each other and produce action. Any such system may be traits which are maintained under constraints exerted by the organisation.

It may be objected that unlike the body systems mentioned above, the mind does not occupy any particular location within the body. Though the nervous system has an especially important role for explaining mental activity, not all of its components and operations directly contribute to the mind's activities. Vice versa, researchers on the embodied mind have argued forcefully that mental activity is constituted by operations of organs other than the nervous system (see Shapiro and Spaulding 2021). For example, the pounding of my heart may literally be a constituent of my fear. Thus, the mind may be co-located with the whole body, rather than occupying a particular location within it.

However, just because a subsystem cannot be exactly located does not imply that it doesn't exist. For a comparison, consider societies. They can be described as complex systems, if not as self-maintaining systems realising closure. Within them, we can describe subsystems such as the political and economic systems or the 'dating economy'. These are distinct, but not easily localisable in space and time. Members of a society may simultaneously be working alongside each other, debating politics and flirting. Thereby, they are participating in different, mutually interacting

subsystems at the same time. The mind as a subsystem may relate to the organism in a similar way as these subsystems relate to society.

It may also be objected that character traits do not correspond to a unified feature of the mind. Rather, a particular character trait, like John's shyness may consist in a mere collection of different features. Each of them explains John's shy (re)action in different circumstances. However, if each of the features exert similar constraints, the organisational account may still treat them as a functional trait. It is maintained, in its actual extent, under all the constraints under which the elements of the collection are maintained. As we will see below in more detail, if the agent or her society has a concept of the character trait in question, the collection may be further unified by how its elements are regulated.

The objector may also argue that character traits simply are not real and that, for that reason, they cannot be functional. They may suggest either that no multipartite model of the mind is apt, or that humans do not exhibit the constant characters that character traits are supposed to explain. Given that virtually all virtue ethicists operate with multipartite models of the mind, and that they appeal to character traits, these are objections to virtue ethics in general rather than to the revision of neo-Aristotelianism I am proposing. Hence, I will not offer a detailed response to this objection.

However, firstly will note that the existence of character traits seems to be well confirmed by experience. Some people are shy, some aren't. Furthermore, later empirical work suggest that there are stable behavioural dispositions. Consider, for example, the so-called *cognitive-affective personality system* (CAPS) model, which focuses on stable tendency to 'encode' certain situations. This model has been defended by Walter Mischel (1973, 2009), whose experiments helped kickstart situationism. The strongest empirical evidence against character traits is in tension only with the wide-spread occurrence of 'global' character traits, such as patience or thoroughgoing cruelty. People who act 'patient' in one situation may not be especially likely to act patient in situations of a different kind. However, this only shows that few people have (near-)perfect virtue.

Regarding the former suggestion, I will point out that multipartite models of the mind are extremely common and show success in explaining and predicting behaviour. If no such models are apt, this would be extremely surprising.

2.2 Character traits contribute to self-maintenance

As we have seen that character traits satisfy C2-3, I will now describe several ways in which character traits can contribute to self-maintenance and satisfy C1. Specifically, they can contribute to the *basic self-maintenance of the organism*, they can contribute to the *self-maintenance of the mental organisation*, they can contribute to the *self-maintenance of lineages and quasi-lineages*.

Character traits are traits of a human organism that reliably produce certain behaviours. Through these behaviours, they affect the organism's environment. Hence, through these behaviours, a character trait may exert constraints under which other traits are maintained. For example, fearfulness or caution may exert the constraint of an absence of injurious external forces, acquisitiveness or foresight the availability of material goods like food. They may also produce circumstances that are presupposed by other traits that affect behaviour. For example, geniality may produce a good reputation, which may be presupposed by other traits such as when I ask my friends to help me move into a new home. Living in a (well-furnished) home increases viability.

Besides relating to physical maintenance in this way, character traits may also contribute to *mental organisation*. It is plausible that the organisation of the mind according to a given multipartite model is not completely determined by the other subsystems of the organism. Rather, the mind in part maintains itself and its own organisation. For example, as described in the theory of Emotion Regulation, humans use mental strategies such as reappraisal, distancing, suppression and distraction to regulate their own emotions. Thus, they prevent themselves from becoming either overwhelmed with emotions or apathetic and unable to act. Plausibly, positively valenced emotions play the role of a 'signal' for the agent to persist in their current strategy since things are going well. In contrast, negative emotions are a signal to change one's behaviour, and, in extreme cases, to trigger thoroughgoing mental reorganization. Thus, an important part of mental self-maintenance consists in attaining positively valenced experiences.

The deployment of such strategies, of course, depends on the current emotional and mental state. An agent who is already completely overwhelmed is unlikely to regulate their emotions down. One who is apathetic is unlikely to regulate them up.

Besides the emotional state, beliefs, memories and cognitive abilities are not merely passively stored by the mind as in a library. After every recall, they are 're-stored' in a potentially changed way, with new associations.. If they are not recalled, they may be lost. Hence, they can be thought of as actively reproducing themselves as part of the mental system.
Hence, the mental organisation is itself a partly self-maintaining system with an organisation, the *mental organisation*. This is further supported by the effect that traumatic events can have on the long-term psychological state of an agent, even if they don't have physiological effects. Such events disrupt the mental organisation. Such a mental organisation of course contributes to the self-maintenance of the organism. With constant emotional breakdowns, or without stable beliefs over time, a human agent's viability is reduced.

Character traits contribute to the mental organisation's self-maintenance both through mental events and through producing action. The first is the case, for example, when they involve the use of certain strategies of emotion regulation, or of the recall of certain memories over others. The latter is the case, for example if it causes me to practise certain skills that involve knowledge or to seek out situations that affect the agent in a certain way.

David Wong (2006) describes *effective agency*, which we can see as a common feature of human mental organisations. Effective agents are able to set themselves goals and to take effective means towards these goals. It is clear that effective agency is an important aspect in the ways in which most humans self-maintain. Wong argues that effective agency needs to be actively maintained. Especially, humans are prone to be mistaken about their actual desires and motivations, and therefore set themselves goals but cannot motivate themselves to take effective means or sabotage themselves. Maintaining effective agency thus depends on a fairly accurate knowledge of one's own motivations. Such a knowledge, however, is mainly acquired through interactions with trusted friends and other familiars. Thus, maintaining such relationships is an important contribution to the mental organisation's self-maintenance and the organism's viability. Character traits plausibly contribute to how one maintains friendships, relationships with family members and lovers etc.

Besides contributing to the self-maintenance of the individual organism, traits can also be functional by contributing to the maintenance of *lineages*. As the originators of the organisational account have recognised, the account must be able to explain *cross-generation functions* (Saborido et al. 2011). If the

only relevant self-maintenance consisted in the maintenance of the individual organism, then, e.g., the entire reproductive system would be functionless. However, in producing offspring and securing its survival, the organisation maintains itself across generations. We can say that it maintains itself as the organisation of a *lineage* rather than as that of an individual organism. Hence, character traits can also be functional by contributing to producing and raising viable offspring.

Benefitting one's own offspring is not the only way in which one can contribute to the reproduction of the lineage. One may also contribute to the self-maintenance of others who share a similar organisation. In non-human nature, this occurs frequently, primarily among individuals that share much genetic information. Thus, traits that contribute to the viability of other human beings, and to the continued existence of the social systems in which humans live, can be functional.

In humans, the mental organisation may be passed on partly independently of other aspects of the organisation. We can teach others our beliefs and our cognitive skills, and we can raise and educate them to have certain character traits, use certain strategies of emotion regulation, and have a certain world view similar to ours. Hence besides the *genetic lineages* described above, which involve the reproduction of overall similar organisations, human beings also have *cultural lineages* which involve the reproduction of similar mental organisations. Character traits that contribute to the self-maintenance of such cultural lineages, or the maintenance of the social systems that make their transmission possible, count as functional.

3. The regulated functions of character traits

So far, we have seen how character traits, and thus virtues, can satisfy the organisational account's core conditions and count as functional. In this section, I argue that character traits have *regulated* functions. In doing so, I will describe what distinguishes character traits from other traits with (organisational) functions.

3.1 Self-control as self-regulation

Human beings have a capacity for what is sometimes described as *self-control* or *necessitation*. We apparently can sometimes 'force ourselves' to act in certain ways because we think we ought to, even if we don't 'feel like it', or are not 'inclined to'. We are acting while experiencing a sense of inner

conflict. Different multipartite models explain this common experience in similar ways. They often suggest that different parts of the mind are in conflict, favouring different actions. One part then succeeds in controlling or circumventing the others in determining action. For Plato, it might be *logos* 'reason' which succeeds in controlling the conflicting impulses of spirit or appetite. For Hume (1740, 3.2.2.26), the silent passion of concern for honour and reputation may motivate me to act virtuously if my other desires favour vicious, e.g. unjust, acts. For Kant (2019 [1785]), it is practical reason controlling inclinations. For Harry Frankfurt (1971), it is higher-order desires controlling lower-order desires. Higher-order desires are desires about whether some other desires should determine action. For modularist cognitive scientists like Mercier and Sperber (2017), it is one module, perhaps the reason-module, controlling other modules.

In these accounts, the part which succeeds in determining action in cases of self-control is typically not active, or not as active, in cases where the other parts favour the same action. The desire for honour and reputation is not what motivates me to be kind to my children if I am feeling a strong natural affection towards them. Even for the thinkers where the controlling part plays a role in motivating required action that is also in line with one's other desires, its role appears more passive. It lies more in allowing the other parts to bring the action about.

Thus, the accounts describe instances of self-regulation. The 'controlling' part is a regulating subsystem. It responds to certain features of the containing system: other parts favouring a certain, regulated-against action. It does so by preventing the effect through which that part would otherwise have determined action. If the action that is produced this way actualises the function of the producing system while the action that is prevented is incompatible with that action and contributes to self-maintenance to a lesser extent, then the produced action actualises a regulated function. More precisely, a character trait that could have produced the action without the activity of the 'controlling' part but didn't do so has the regulated function to produce that action. It further has that regulated function if the preferred action is not produced, but the controlling instance is responding to this failure. Finally, it has this regulated function if it is functional and actually produces the action in question, but the controlling part is such that it would have responded if the regulated part had failed.

3.2 Virtue-inquiry as a general regulatory subsystem

The foregoing is not enough to explain what differentiates the objects of ethical evaluation from the objects of other evaluations in terms of regulated functions. Most malfunctions are not ethical failures. Foot and Hursthouse, answer this question by saying that virtues and vices are excellences and defects *of practical reason* or *rationality*, which are sui generis concepts. Under the organisational account, too, virtues are the traits that constitute well-functioning practical reason. However, the notion of *practical reason* is not used in scientific biology. Hence, in order to avoid the challenge from science, the organisational account of virtue must show what differentiates virtues from other functional traits without referring to practical reason. Thereby, it shows that appeals to practical reason aren't superstitious.

I do so by arguing that character traits --- and thus the actions they produce --- are regulated in a particular way. They are regulated by ethical inquiry, which is a *general regulatory subsystem*. They aim at self-maintenance *as such*. Other regulatory subsystems always aim at self-maintenance on some particular level. In this section, I focus on how ethical inquiry's status as a general regulatory subsystem sets the functions it regulates for, viz. virtues, apart from other organisational functions. In Chapter Five, Sections 2.2.2--3, I describe how ethical inquiry's function as a general regulatory subsystem shapes its internal, autonomous logic. In Chapter Seven, I will show it can explain the close connection between ethical inquiry and action, i.e. the capacity of ethical inquiry to regulate behaviour. In this chapter, I take it for granted that ethical inquiry has this capacity.

Most humans engage in *ethical inquiry*. They formulate arguments about how they should act in a particular case, or which character traits they ought to have. They might do so publicly, together with others, or privately in their mind. Agents may act on the conclusion of ethical inquiry, including virtue-inquiry. Sometimes, this may be accompanied by no sense of inner conflict. Occasionally, however, it constitutes a case of self-control, and hence of regulation.

Ethical inquiry can regulate many behaviour-producing subsystems. Most of my movements on a timescale over a few seconds during my waking time are produced by systems potentially regulated by ethical inquiry. It can also regulate other regulatory subsystems that in turn regulate the lower-order behaviour-producing subsystems. For example, as an outcome of ethical inquiry, I may

practice paying more attention to others' vulnerability. This in turn may counteract my tendency towards callousness.

With regard to those behaviour-producing systems which it regulates, ethical inquiry can in principle respond to any failure to exert a given constraint. Thus, it can be functional through responding to any failure to contribute to self-maintenance. Even only looking at the conditions, ethical inquiry is *actually* disposed to respond to a particular human agent, they are likely to be highly diverse. The conditions among them that indicate a failure to exert some presupposed constraint will indicate failures of many different kinds. Thus, we can call ethical inquiry a *general* regulatory subsystem. This appears to be a unique development of humans. Other regulatory subsystems occurring in living systems only respond to a fairly unified set of conditions, and they could not easily respond to circumstances that are radically different.

In the (neo-)Aristotelians' language, virtue-inquiry is aimed at the organism's good *as such*, not only at some particular good for the organism. In this slogan, a *good for the organism* is a constraint that contributes to self-maintenance. This is importantly different from saying that self-maintenance is the only good for the organism. If food contributes to self-maintenance, then food is a good for the organism. If friendship contributes to self-maintenance, then it is a good for the organism. Assuming that an action being good for the organism means that there is a reason for the organism to perform this action, we can further say that ethical inquiry responds to reasons.

Aiming at the good for the organism as such means aiming at some collection of constraints that maximises self-maintenance. Even if we only consider the individual organism, there may be multiple such constraints. As we have seen above, however, besides the individual organism, traits may also be functional by contributing to the self-maintenance of cultural and genetic lineages. Such contributions are likely incommensurable. Hence, there will be different incommensurably optimal collections of constraints which contain different trade-offs. Aiming at the good of the organism as such means aiming at realising some such collection *de re* or *de dicto*. Thus, ethical inquiry differs substantially from other regulatory subsystems. This explains how ethical evaluations can be substantially different from other evaluations in terms of organisational functions.

3.3 Character as dispositions regulated by ethical inquiry

The collection of dispositions which is regulated through ethical inquiry is unified through this fact. This collection contains ethical inquiry itself, since as we have seen, ethical inquiry can indirectly regulate itself. We can call this collection the *character* of the agent. The character is constituted by the character traits. The behaviour that is produced through the character are *actions*. Thus, the bursts of coprolalia associated with Tourette's syndrome are not actions, since they cannot directly be regulated by ethical inquiry. At most, they can be regulated through other actions.

The regulated functions of the elements of character derive from their regulation through ethical inquiry. They are partly shaped through ethical inquiry. Hence, we can say that like ethical inquiry, character aims at the good of the organism as such. Hence, character counts as responding to reasons just like ethical inquiry. Thus, practical rationality, or the ability to respond to reasons, is not just a matter of deliberation and self-control. Rather, it is a feature of the whole character. It involves attentional, perceptual and emotional dispositions as much as deliberation. In fact, ethical inquiry may be causally rather irrelevant to most actions. It is simply the possibility of regulation through ethical inquiry that transforms the normative status of character traits.

This explains a disparity between our evaluative talk about non-human living beings and about humans which other virtue ethicists, such as Hursthouse (1999), fail to explain. With human beings, we call traits good (or bad), and the whole individual. We say that physical strength is a good bodily trait, and kindness is a good character trait. Yet, we also say that Georg Elser was a good human being. In contrast, with non-human organisms, we ordinarily only use the first type of ascription. We say that a particular eagle's eyesight is good or that an oak has good roots if it has strong ones. But we would not ordinarily say that they are a good eagle/oak. I suggest that this is because with human beings, the name or description of the whole individual may stand, *totum pro parte,* for the character. Thus, saying that Elser was a good human being is a way to say that he had a good character. Such synecdoche is possible because our characters determine much of our behaviour. They are of central importance for us in interacting with each other. In contrast, with other living beings, there is no comparable system. Hence, it makes no sense to speak of a *good eagle* even though it makes sense to speak of an eagle's good talons.

Hursthouse (1999) uses phrases like *good specimen/exemplar of the eagle* as more natural-sounding paraphrases of *good eagle*. However, notions like *specimen* and *exemplar* reflect human concerns in studying and categorising organisms. It is in part a coincidence that overall well-functioning specimens are generally most useful in those practices. Even the most serious physical diseases are irrelevant to our assessment that someone is a *good human being*. In contrast, an eagle with late-stage cancer would hardly be a good specimen.

Additionally, it explains why the distinction between 'actions', or what an organism does, and reflex behaviour is more clear-cut with humans than with other animals. For example, we say that a dog raises its hackles and a porcupine its quills. The human version of that phenomenon --- piloerection --- is something that merely happens to us. We *get* goosebumps. Being incapable of ethical inquiry and not having a character, the distinction between reflex-like behaviour and action in dogs and porcupines is much more vague than in humans.

3.4 Summing up

To sum up, there is a difference between the regulated functions of character traits and other regulated functions. This explains why ethical evaluation differs from other sorts of evaluation in terms of organisational functions. It also goes some way towards explaining why ethical evaluations involve criticism and blame of the agent in a way other evaluations do not. Firstly, as evaluations of character traits, they are evaluations of the agent's character which those traits constitute. The character can be identified with the agent themselves. Hence, unlike other evaluations they are evaluations of the agent themselves.

Secondly, agents are responsible for their characters in a way in which they are not for other traits. As we have seen, an agent's mental organisation is partly self-maintaining. The same is true of the character, which is a subsystem of the mental organisation. My character traits, for example, reproduce themselves through repeated practice. I, identified with my character, could have changed my character in a way in which I could not have changed other aspects of my organisation, such as the operation of my heart. Of course, in a sense I am not ultimately responsible for my character. My character traits were originally shaped by my upbringing, and there are likely influences on my character from other aspects of the organism that are not under the control of character. However,

there is still a level of control over character traits that doesn't exist with regard to other traits with regulated function.

Finally and relatedly, blame and praise can itself influence character. Memories or anticipations of blame and praise may motivate us to act in certain ways. Decoupled from actual praise and criticism, the notions featured in moral evaluations can also feature in ethical inquiry. I may exercise self-control in not performing a certain action because it would be *unjust* or *not what I ought to do*. Thus, moral criticism and praise can constitute a kind of encouragement or 'call to action' in a way other evaluations based on functions do not. I will discuss this *social* nature of ethical inquiry later, in more detail.

4. Regulated functions and the received virtues

In the preceding sections, I have argued that character traits generally have organisational functions, on which ethical evaluations can be based. In this section, I will begin to argue that it is at least possible that virtues according to received commonsense (*received virtues* for short) --- e.g. charity, moderation, justice, etc. --- have such regulated functions. I do not want to build into the organisational account of virtue the conservative assumption that these notions of virtue are accurate. However, if it could not accommodate them, this would open my account up to similar objections as the one presented in chapter one against accounts identifying virtue with a kind of selected-effect function. Then, whatever virtue is, given our strongest beliefs about it, it could not be a kind of organisational function. I will also show how certain revisionist proposals, such as ones suggested by the work of Peter Singer (e.g. 1981), can in principle be accommodated within the framework.

In this section, I will primarily argue that the received virtues are functional, i.e. that they contribute to self-maintenance. Thus, in an agent whose ethical inquiry does in fact tend to regulate for them, they will have regulated functions. Thus, according to the account I am defending, in such an agent they are virtues. Their failures are malfunctions, and character traits that imply their failure are vices. This is sufficient for a 'Humean' account of individual virtue, resembling Street's (2012) 'Humean constructivism'.

In Chapter Five, I will describe well-functioning ethical inquiry in more detail. My discussion will imply that for many agents whose ethical inquiry superficially fails to regulate for received virtues, it would do so if it maximally actualised its function. As I will argue, this is enough for these to count as regulated for. In Chapter Six, I argue that certain virtues are genuinely universal: *all* ethical inquiry maximally actualising its function regulates for them.

We can distinguish different categories of received virtues based on whose benefit they are aimed at. Firstly, there are self-regarding virtues like moderation, temperance, prudence and courage insofar as it applies to the pursuit of personal ends. With these virtues, the primary beneficiary is the virtuous agent themselves. Then, there are other-regarding virtues. Examples include charity, kindness and justice. Here, others are the primary beneficiaries. Among the other-regarding virtues, we can distinguish partial virtues and impartial virtues. Partial virtues benefit those who are in special relationships to the agent: their kith and kin; their friends, family members and fellow citizens. Generosity plausibly is a partial value. Impartial virtues benefit others independently of their relationship to the agent. They apply to strangers as much as to friends and lovers. Justice and charity to needy strangers are potential examples.

Many received virtues, such as courage or charity, have aspects that fall into more than one of these categories. For simplicity's sake, however, I will discuss virtues as if they fall neatly into one of these categories.

In the following, I will first discuss the functions of self-regarding and partial virtues. Then, I will discuss impartial virtues. Finally, I will discuss an objection concerning the influence of biases on regulatory subsystems.

4.1 Self-regarding and partial virtues

It is easiest to see how self-regarding virtues have organisational functions. To be moderate in eating, exercise, drink etc. clearly contributes to your self-maintenance. The same is true of prudent planning, which may prevent me from falling into misery. Even if self-regarding virtues do not so directly contribute to the organism's basic self-maintenance, they may enable them to achieve their

goals effectively. Achieving one's goals through effective action is an indispensable part of most people's self-maintenance.

However, this seems to imply that all self-regarding virtues maximise viability, and hence longevity. However, seeking some pleasures and enjoyments that may reduce your viability can be an expression of some received self-regarding virtues. Enjoying wine and rich food in pleasant company may be more virtuous than pursuing good health above all else. The agent adopting the latter strategy may seem like a hoarder of life-time. They resemble the miser who amasses money but never spends it on anything valuable. It's as much a failure of moderation as over-indulgence.

To see how moderate enjoyment can be virtuous, we consider its relation to mental organisation and effective agency. Plausibly, effective agency is most likely to be maintained if the agent experiences a fair amount of pleasure. Hence, excessive self-denial risks the breakdown of effective agency. An agent may fall into despondency or swing into reckless abandon. Even if the self-denial does not decrease the viability of the organism, it may decrease the viability of the *mental* organisation. Hence, there is a trade-off between the basic self-maintenance of the organism, and of the mental organisation. This trade-off allows for a variety of notions of the 'good life' between asceticism and enjoyment.

Some humans appear to be contented ascetics. They reap the benefits of self-denial without the downsides. Perhaps they shaped their desires so that living in a way that maximises longevity but doesn't involve more self-denial than other human lives. Or, they may have unusual 'strength of character', which allows them to maintain their mental organisation despite extraordinary self-denial. However, it is plausible that living such a life takes unusual luck. Externally, the contented ascetic may have to be in a particular kind of environment. Internally, they must have a particular kind of constitution.

Regarding external conditions, an organism's viability is determined with regard to a set of counterfactual circumstances. Otherwise, the degree of functionality of an organism's component could not be distinguished from the actual length of its life. Hence, a trait does not fail to count as functional because in the precise actual circumstance it did not contribute optimally to self-maintenance if it could have easily done, and likely would have done so if things were any

different. This introduces another set of trade-offs: self-maintenance in narrow actual circumstances versus self-maintenance in a broader range of circumstances that could more or less easily have been the case.

With regard to internal circumstances, this shows that we should be cautious in criticising agents as vicious for not being ascetics. They may lack the constitution to become content ascetics. This may be unrelated to their character, in which case they are not responsible for it. In fact, even if an agent were able to transform themselves into a contented ascetic, this may involve such fundamental changes in their mental organisation that it would not count as self-*maintenance* of that organisation.

The relation between enjoyment and self-maintenance of the mental organisation is not coincidental. We may think of enjoyment as consisting of positively valenced, nice feelings are the experienced satisfaction of one's desires. In either case, it involved the neuro-psychological 'reward system'. Evolutionary, this system has been selected to reinforce behaviours that contributed to the individual's survival. Throughout most of human evolution, if the reward system was engaged little over a long-time, this meant that the current behaviour was likely not working in securing survival. Hence, it would be more likely to promote survival to change behaviours, i.e. to adopt a new mental organisation. In our current environment, low engagement of the reward system does not indicate a lifestyle unconducive to survival. For example, foregoing opportunities to consume high-calorie foods may benefit viability, whereas it would have been unlikely to do so in evolutionary history. Such evolutionary preferences are of course always shaped and transformed by culture. Nonetheless, we need to find a balance between actions that optimally contribute to our basic viability, and ones that engage our reward system and enable self-maintenance of the mental organisation. Thus, for beings like us, who have complex mental lives and live in conditions different from those they evolved, it is no coincidence that enjoyment is an additional aim of virtues that may be in conflict with basic viability. If we lived the same way as our earliest *homo sapiens* ancestors, we would have no need for temperance, moderation --- or self-care.

Partial virtues are involved in maintaining close relationships, which benefit our self-maintenance. Our familiars can give us physical assistance. They may feature in our emotional self-regulation. Importantly, they may help us to gain knowledge of our motivations. Thus, they help us maintain effective agency and avoid self-sabotage. Saying that these contributions to self-maintenance ground

the functionality of other-regarding virtues does not mean that agents must value those contributions for them to have regulated functions. If friendship is a constraint that contributes to self-maintenance, and an agent values friendships for their own sake, then character traits contributing to friendships will have regulated functions. In fact, valuing friendships for their contributions to self-maintenance might diminish one's ability to maintain friendships and thus to enjoy those contributions.

Besides such contribution to individual self-maintenance, other-regarding virtues also contribute to self-maintenance of lineages. With regard to virtues regarding family members, especially one's children, these are genetic lineages. Regarding others, such as friends whose worldview one shares or members of the same cultural tradition as oneself, these are cultural lineages.

4.2 Impartial virtues and quasi-lineages

The functionality of impartial virtues may appear to be more dubious. After all, there is no guarantee that helping a stranger will benefit me. Furthermore, if the virtues depended on the beneficiary sharing a cultural or genetic lineage with me, they would not be impartial.

However, I argue that there are *quasi-lineages* which are shared by large groups of creatures. Their historical connection may lie in the deep past, or it may only be brought about in benefitting other members of the lineage. To see how contributions to the self-maintenance of quasi-lineages are functional, consider again the motivation for introducing the concept of lineages into the organisational account of function. It is needed to account for cross-generation functions, such as reproduction. Traits that contribute only to the production of offspring or to the offspring's viability are functional because of that. They contribute to the self-maintenance of the organisation in other organisms. Thus, they are contributing to the self-maintenance of the organisation in a lineage. However, some animals also show all kinds of other seemingly functional altruistic behaviours. For example, they alloparent. They raise (increase the viability of) their relatives' offspring. Given that the beneficiaries of such altruism have a very similar organisation, this can also be seen as contributing to the self-maintenance of a lineage. However, the organisation of the beneficiary is never *exactly* the same as that of the benefactor. Hence, there is only a *gradual* difference between one human benefitting the viability of their daughter, their nephew, a complete stranger's son---or a non-human fellow mammal. Traits producing such actions are all at least minimally functional.

One such quasi-lineage is the human lineage. All humans clearly resemble each other in many organisational ways. They tend to have the same body plan, the same organs; they tend to have similar psychologies, featuring beliefs and desires. Thus, if a trait of a human being benefits the viability of any other human being, it contributes to the self-maintenance of the deep human lineage. This can ground the functionality of impartial virtues applying to all humans. This includes the received notion of justice as well as a radical altruism resembling the proposals of Singer (1981). In addition, there is a *sentient quasi-lineage*. My organisation does not only resemble that of other humans, but also non-human sentient animals. Hence, kindness and charity towards animals (as well as a radical anti-speciesist altruism) can be functional.

One might worry that this links our duties to distant strangers too much to our similarity to them. However, it does not imply that our obligations to help strangers are somehow proportional to how similar they are to us. They are all equally members of the relevant quasi-lineage. Strangers that are more similar to us may additionally share some shallower lineages with us. However, contributions to the deeper and the shallower lineages can often be incommensurable. Incompatible traits favouring the former or the latter could both be functional. Which one constitutes a malfunction depends on the regulatory subsystems, i.e. on the agent's values. It also does not imply that the agent must value the contribution to the lineage..

Sceptics may point out that functional traits contributing to quasi-lineage seem not to occur in non-human species. Does the appeal to quasi-lineages violate the continuity between virtue and biological function? However, the reason that such traits rarely if ever occur in non-human species is that they tend not to be selected for. A trait benefiting a quasi-lineage is likely to benefit organisms that do not share this trait. Thus, they are unlikely to emerge through natural selection. In human beings, such traits could likely emerge because technological progress has lifted some selection pressures. Furthermore, they can be side-effects of traits that enable the exceptional general purpose reasoning of humans and their ability to live together in large, diverse groups. The notion of quasi-lineages follows naturally from the organisational approach to cross-generation functions. The organisational account deliberately excludes evolutionary considerations from the definition of *function*. Hence, the fact that traits benefitting fellow members of quasi-lineages haven't been selected to so does not show that those effects are not functional.

5. Conclusion

In the preceding sections, I have argued that character traits have regulated functions. To show this, I have argued that multipartite models of the mind describe a subsystem of the human organism on a high level of description. The functions of character traits are of an unusual kind, which can ground ethical evaluations. In showing this, I have explained the notion of character within my framework. I have further shown that the account is compatible with the received virtue virtues being actual virtues. In the next chapter (Chapter Five), I will focus on the nature of ethical inquiry according to this account. Thus, I will develop the organisational account of virtue into bio-constructivism, a more complete account of ethical discourse. This will show that the received virtues are not just *possibly* virtuous. Rather, at least for us and our communities, they are *actually* virtues. In the following chapter (Chapter Six), I will further argue that at least some of the received virtues are virtues for all human beings.

FIVE: … to bio-constructivism

Ethical inquiry according to the organisational account of virtue

In this chapter, I will describe how ethical inquiry is possible according to the organisational account of function. Thus, I will move on from the mere organisational account of virtue to bio-constructivism: an account of ethical judgement and ethical truth in general rather than merely of virtue. I will first describe the phenomenon of ethical inquiry. Then I will provide an apparent dilemma for the organisational account of function. Seemingly, it must either be based on a version of the organisational account of function that makes different levels of self-maintenance commensurable, or it is based on one under which they are incommensurable. On the first horn, the account of virtue would have to 'hypostatise' a way of being human (cf. Chapter Two). It would have to privilege it in a way that is not supported by the appropriate reasons, and remove it from any ethical criticism. Consequently, distinctively or 'autonomously' ethical inquiry would be impossible. Ethical inquiry would turn out to be merely a branch of scientific inquiry. On the other hand, if the organisational account of virtue is based on an account of function according to which different levels of self-maintenance are incommensurable, then 'ethical inquiry' would be trivial. With few constraints, whatever conclusion an agent reaches would thereby be true.

In order to resolve that dilemma, I will provide my positive account of virtue-inquiry and its peculiar relation to its subject matter. *Virtue-inquiry* forms a part of ethical inquiry. In virtue-inquiry, an agent inquires not simply about what to do, but about which considerations to treat as reasons for what actions. Thus, they inquire about which shape the remainder of ethical inquiry should take and which character traits, or dispositions to treat considerations as reasons, are virtues. It is because of virtue-inquiry that ethical inquiry is a general regulatory subsystem. Virtue-inquiry regulates other subsystems, and through its regulation it systematically contributes to self-maintenance *in general*, rather than to self-maintenance on a particular level. As such, it plays a central role in determining the function of the traits it regulations. As long as they make a contribution to self-maintenance on *some* level and are regulated for by virtue-inquiry, they are functional, and thus virtuous. However, this does not happen if virtue-inquiry fails to actualise its own function in (seemingly) 'regulating' for them. This would not be genuine regulation under the organisational account. Hence, in virtue-inquiry the inquirer simultaneously attempts to ensure that her virtue-inquiry fully actualises

its function, and thereby that it accurately describes its subject matter, viz. the organisational functions of her character traits. Given that these functions are as much a feature of reality as any other function, bio-constructivism avoids the challenge from science. Successful virtue-inquiry describes reality as much as any other successful inquiry into organisational functions. However, virtue-inquiry, and thus ethical inquiry as a whole, follows its own distinctive logic. This logic is determined by its function rather than by the methods of science. Hence, it avoids hypostatisation.

First, I will argue that the hypostatisation objection and the naturalist response to the challenge from science together entail a form of relativism. Which traits count as virtues depends on considerations that are local to the agent whose traits they are. Given this, I will argue that it is plausible that it depends on the agent's prior ethical judgements. Thus, it is also plausible that ethical inquiry proceeds on the basis of such prior ethical judgements, similar to the 'holistic' method of Neurath's boat described for ethics by Hursthouse (1999). We discover which ethical judgements are true of us on the basis of ethical judgements we find ourselves with. In other words, we revise our ethical judgements. I will describe a number of constraints on a virtuous or reasonable agent. These constraints correspond to and justify 'moves' in ethical inquiry. Together, these moves constitute the distinctive logic of ethical inquiry.

Finally, I argue that maximally functional virtue-inquiry, and thus ethical inquiry, will (almost) completely converge *within* an ethical community. This is due to the function it has in enabling cooperation. I continue discussing this topic in Chapter Six, where I focus on evaluating distant communities and arguing for universal virtues on which *all* maximally functional virtue-inquiry would converge.

1. The phenomena of ethical inquiry

Ethical inquiry is a form of inquiry, i.e. an effortful attempt to form or revise some judgements or answer some questions. An inquirer has the aim of 'getting it right', or avoiding 'mistakes', and invests time and energy in achieving that aim. She is not merely 'picking' which judgements to endorse. Ethical inquiry, which in my usage includes prudential inquiry, is distinguished from other forms of inquiry in that it is a form of normative inquiry about action. It is about what one ought to do, should do, must do, what is right, appropriate or fitting to do etc. In many cases the questions

are not explicitly couched in such 'thin' moral terms. Rather, they feature 'thick' ethical concepts which have some descriptive content, but also a conceptual connection to thin moral terms. These thick moral terms typically designate either action types or character traits. Thus, ethical inquiry often concerns questions such as 'what is courageous/cowardly/craven/cruel, etc'. A courageous person is a person who responds appropriately to danger, i.e. as one ought to. A courageous act is one that is typical or characteristic of a courageous person.

Ethical inquiry is further distinguished from other kinds of normative inquiry. It is distinguished from epistemic inquiry in that its conclusions are about actions and motivations rather than about judgements or beliefs. It is distinguished from inquiries concerned with skill in that its judgements apply to agents independently of any particular, contingent aim. An inquiry concerned with skill might lead to the conclusion, say, that 'one ought to put pasta into water that is already boiling rather than heat it up with the water'. However, I can legitimately say 'it is not true that I ought to do that; I am not trying to cook a pleasing dish, I am trying to cook a mediocre dish for my archnemesis!' Such moves are not possible with conclusions in ethical reasoning.

Ethical inquiry is also distinct from other inquiries, such as those concerned with the rules of etiquette or of a game by its connection to action. This connection with action distinguishes it, for example, from inquiry into rules of the games. Inquiry into the rules of a game may also produce categorical imperatives. For example, it is true, as a fact of the rules of chess, that one must move the king out of check or end the check in some other way if the king is in check. This is true even if the player in check does not want to play chess anymore and has no other aims that could be achieved by continuing the game. However, unlike judgements in ethical inquiry, such judgements have no strong connection to action. Perhaps, an agent who recognises that the rules of chess require her to move the king out of check might indeed move the king out of habit. However, if it also occurs to her that she doesn't have any aims that are furthered by playing chess, this connection will be disrupted and she will not keep playing chess. In contrast, if I reach the judgement in ethical inquiry that courage requires me to save a drowning child, then this judgement is capable of motivating me to save the drowning child. It could do so even if I also judge that saving the child serves no independent aim of mine and interferes with other aims I have. Unlike in the case of judgements about the rules of a game, such a judgement does not disrupt the connection between the normative

judgement in action. This does not mean that judgements in ethical inquiry necessarily lead to action. There may, for example, be weakness of will.

Some ethical inquiry is concerned with particular actions. For example, I may inquire about whether to compromise my policy of not eating meat for environmental reasons in order to be polite to an older member of my extended family. However, much ethically motivated action does not involve ethical inquiry. If I see a child in danger of falling off a cliff and saving the child does not involve any significant risk of harm to me, I will most likely not inquire into whether to save the child. I will simply respond by grabbing the child to save it. However, an agent may engage in ethical inquiry concerning a particular action if there are competing considerations of similar strength.

A lot of ethical inquiry simply instantiates the inquirer's character traits, i.e. their disposition to treat considerations of reasons. If I inquire into whether to save the drowning child despite danger to me, and I am virtuous, then my inquiry will instantiate my beneficence and prudent caution. I will simply reason 'There is some danger to me if I try saving that child. However, the danger is quite small and the child dying would be terrible. Thus, I must save the child'. Such reasoning instantiates my disposition to treat considerations of my own safety and of the need of others as reasons. If I am not virtuous, it may instantiate my callousness and my cowardice. This is also the case for some ethical inquiry concerned with character traits. For example, the inquiry in the conclusion that I should be more daring in order to be courageous may simply instantiate my (imperfect) courage. Looking back, I judge that I was too timid in situations where I should have been more daring. In doing so, I treat certain considerations that were relevant at the time as reasons, even though I failed to treat them as (sufficiently strong) reasons then. In doing so, I exhibit the virtues of *charity* and *prudence* without referring to either of them.

Sometimes, however, I inquire into which considerations to treat as reasons. This inquiry is often explicitly about the virtues: I may ask myself what kind of actions are prudent, charitable, courageous, etc. I may also ask myself whether to act charitably at all, etc., i.e. whether charity really is a virtue. This is *virtue-inquiry*. Such inquiry may be occasioned by a conflict between two apparent virtues that the agent can't resolve in the simple way described above. I may be unsure which consideration to weigh more strongly and begin to reflect about the nature of the virtues involved.

Virtue-inquiry may also be occasioned by more general practical doubt about my character and my way of life, or by simple curiosity.

Agents can conduct such higher-order ethical inquiry into which considerations to treat as reasons without appealing to the notion of *virtue* or particular virtues. This is the case, for example, with inquiry into whether Kantian deontology or utilitarianism is correct. However, we can subsume such inquiry into virtue-inquiry. It can be seen as the inquiry into a particular virtue, viz. beneficence or justice, and whether all ethical or moral requirements can be derived from this virtue.

In Chapter Seven, I will provide an explanation of the link between ethical judgements and action based in the organisational account. In this and the next chapter, I will take the link for granted and focus on the logic of ethical inquiry. I will also begin by focussing on self-directed ethical inquiry, i.e. inquiry in which an agent attempts to answer questions about how they ought to act or what kind of person they ought to be. In Section 4, I will focus on how we can evaluate others in our own communities and those we interact with. In Chapter Six, I will describe how we can evaluate members of distant communities and support universal ethical judgements.

Besides its connection to action, another important feature of ethical inquiry is what is sometimes called the "autonomy of ethics" (e.g. Prior 1960, Lott 2020). In ethical inquiry, ethical judgements must be supported by other ethical judgements. Ethical inquiry is sui generis, and proceeds by its own methods. It is not simply a branch of any (other) special science, such as biology. This is closely related to the problem of hypostatisation described in Chapter Two. With any conclusion that is reached in a form of inquiry that is not ethical inquiry, we can always challenge its pertinence to action through ethical critique. We can always inquire into the question: '... but is that really how one ought to act?', e.g. into the question: 'Biological inquiry shows that not having children before the age of 25 is dysfunctional, but does that mean that one ought to have children before age 25?'.

2. The apparent dilemma between hypostatisation and trivialisation

In this subsection, I argue that the organisational account of virtue does not entail that there is a 'science of ethics'. It does not make ethical inquiry into a particular form of scientific inquiry. I show this by considering a dilemma between two slightly different attempts to use the organisational

account of virtue to found scientific inquiry into ethical truth. The first is based on accounts of function that make different levels of self-maintenance commensurable, e.g. by privileging one level of self-maintenance. Such accounts lead to unacceptable implications and hypostatisation. The second leaves levels of self-maintenance incommensurable. It trivialises ethical inquiry. Nearly every conclusion reached by an inquirer would be *ipso facto* correct, so there would be no point in expending effort on ethical inquiry. Instead, in the next section, I will argue that the organisational account of virtue supports an account of ethical inquiry as proceeding by its own logic, distinct from scientific inquiry, while still representing facts and the world accurately in the very same sense as the natural sciences if successful.

In the preceding chapter, I have introduced the Organisational Account of Virtue. This account can meet the challenge from science introduced in Chapter Two by showing that the truth and falsity of ethical judgements can be explained by natural science, namely by biology. A trait of a human agent's character is a virtue iff it is functional according to the organisational account of function. It is functional if it contributes to self-maintenance and is regulated for. Thus, the truth or falsity of an ethical statement may play a role in a scientific explanation of why we make the judgement in question. Ethical inquiry is not merely superstitious.

I have also argued that, with human beings, self-maintenance operates on various different levels. This is not a uniquely human feature. However, with humans it is more intense and there are more levels of self-maintenance. Human beings maintain, firstly, on the level of the body or organism. However, on a more macro level, self-maintenance also occurs on the level of lineages, as in procreation and kin-altruism as well as 'quasi-lineages'. On a more micro level, human self-maintenance occurs in the 'mental system'. The human mind is itself a self-maintaining system. This self-maintaining system can collapse even while the organism survives, in a (complete) mental breakdown. Pleasure or positively valenced emotions in general, play an important role in mental self-maintenance, as they reinforce an existing mental organisation.

Given this, a dilemma arises from how the different levels of self-maintenance are to be weighed against each other. Suppose a trait makes an optimal contribution to self-maintenance on one level, but it is incompatible with traits that make greater contributions to self-maintenance on a different level. This situation is ubiquitous in human life: A genial disposition to drink wine with my friends

may contribute to my mental self-maintenance as well as more broadly to my ability to achieve my goals (with the help of friends) and thus to benefit the self-maintenance of my lineages and quasi-lineages. Yet, it also likely shortens my life-expectancy, interfering with self-maintenance on the level of the individual organism. Similar points apply to classical debates concerning conflicts between the apparent virtues of prudence, generosity and charity.

Which among competing dispositions --- e.g. social drinking vs. teetotalism --- is functional depends on two factors: Which, if any, makes the greater overall contribution to self-maintenance, and which is regulated for. The latter factor is strictly secondary to the former. If one trait makes a lesser contribution to self-maintenance overall, i.e. if having that trait rather than the alternative trait inhibits self-maintenance, then no amount of apparent regulation could make it functional. Regulation is defined as a response to the failure of a contribution to self-maintenance presupposed by other systems. Thus, if the trait doesn't contribute to overall self-maintenance, then it cannot be regulated for. The definition is not merely an arbitrary stipulation. Rather it is necessary to distinguish regulation from other kinds of processes in which a subsystem responds to the absence of a constraint in the system.

The organisational account of virtue must endorse one of two options: either the organisational account of virtue is based on an account of function which entails a method of weighing the different levels of self-maintenance, i.e. which makes contributions on different levels commensurable, or it is based on an account which makes them incommensurable. Both options seemingly entail unacceptable consequences for the organisational account of virtue.

2.1 Levels of self-maintenance are commensurable

Firstly, let's assume that the organisational account of virtue is based on an account of function that entails a method of weighing different contributions against each other. Thus, these contributions are commensurable as far as that account of function is concerned. I will argue that such an account would be guilty of hypostatisation in the same way as etiological and propensity-based accounts (see Chapter Two). I will show this, first, by considering particular, plausible methods of weighing different contributions to self-maintenance. Then, I will argue that an account of virtue based on such accounts of function would generally be 'hypostatising' in an objectionable way. It would

privilege a certain way of being human without providing reasons for it within ethical inquiry and while precluding any possibility of questioning or critiquing it within ethical inquiry.

Plausibly, there are several closely related but distinct notions of function which can all be captured by versions of the organisational account of function, and which each have a legitimate role in scientific practice. These different accounts differ by *privileging* one level or another of self-maintenance. For example, an evolutionary biologist interested in the development of populations over timespans much longer than the lives of individual organisms may rely on a notion of function that privileges the maintenance of the lineage. A medical scientist may rely on a notion of function that privileges the level of the individual organism, and a psychiatric researcher that of mental self-maintenance.

By *privileging*, I mean that all contributions to self-maintenance are reduced to contributions on the privileged level. The privileged level is made the 'common currency' of self-maintenance. Contributions on the non-privileged levels are only considered insofar as they indirectly contribute to the privileged level. After all, the different levels of self-maintenance are interdependent. An organism cannot reproduce and maintain the lineage if it cannot survive at least for some time. Survival, on the other hand, can depend on the maintenance of the lineage, as when human children support their elderly parents, or if a buffalo herd consisting of multiple generations fights predators. While such accounts of function have a legitimate role within the life-sciences, accounts of virtue based on them could not explain ethical inquiry.

2.1.1 'Health freaks' and accounts privileging the individual organism

If the privileged level is that of the individual organism, the uniquely most virtuous human agent would resemble a 'health freak': someone who pursues their own health over everything else. For example, they would likely be tee-totallers and abstain even from social drinking. They would also likely invest significant resources in maintaining their own health, attending expensive gyms and consuming expensive but healthy foods. In fact, if they live in a modern industrialised society they would likely pursue a high-earning career and invest their resources into maintaining their health. Statistics show that high earners have a higher life expectancies, and at least some of this is likely due to causation (see e.g. Freeman et al. 2020).

Of course, this does not mean that the virtuous agent would actually only value their own longevity. As with other goods, such a stance is likely self-undermining. Friendships contribute to longevity, and it is easier to maintain friendships if one values the welfare of one's friends in a non-instrumental way. However, a virtuous person under such a 'longevity-first' organisational account of virtue would nonetheless often make choices primarily on the basis of its contribution to health. For example, the 'health freak' might make some apparent sacrifices in most of his life as part of a strategy of reciprocal altruism. For example, they may invest in raising children as their children will then be disposed to look after them in old age. However, they would become less and less likely to make such sacrifices the older they become, as there will be less opportunities to receive rewards in the rest of their life. The 'health freak' may not be an utterly reprehensible character, but it seems unacceptable that their way of life should be the uniquely virtuous way of life for human beings.

2.1.2 'Brood parents' and accounts privileging the lineage

Instead of the level of the individual organism, an account of function may also privilege the level of the lineage. The virtuous agent according to such an account may be called, with some exaggeration, a 'brood parent'. The 'brood parent' would have character traits which maximises the number and offspring they leave, and the number of offspring that their offspring leaves. As with the 'health freak', this doesn't mean that they only value their lineage. Reciprocal and kin-altruism contribute to the self-maintenance of the lineage, and an agent may better secure the survival of their lineage if they treat such altruistic ends as non-instrumentally valuable. However, an agent with character traits that optimise the likelihood that their lineage survives indefinitely would almost certainly have much more offspring than is currently common in industrialised countries. They would have large numbers of offspring, as more offspring minimises the chances that the descendants die early and the lineage becomes extinct. In fact, it seems plausible that a (male) 'brood parent' who is unable to secure mates otherwise would be disposed to resort to sexual violence.

2.1.3 Accounts privileging the 'species'

One might also suggest that the level of the 'species' is the privileged level of self-maintenance. This suggestion is problematic from a biological point of view, as the notion of a species is much vaguer

than the levels of the lineage or organisms. Thus, one might worry that it cannot be made precise enough to support an organisational account of function.

Be that as it may, an account of virtue based on an organisational account of function that privileges the 'species level' would be similarly unacceptable. Admittedly, one might argue that, in general, the best way one can contribute to the maintenance of the species is to look after one's own maintenance and the maintenance of conspecifics close to one. However, once the agent who is virtuous according to this has reached an age where they will not reproduce again, it would be vicious for them to act in any ways that advance individual self-maintenance over the self-maintenance of younger conspecifics. After all, the maintenance of the species beyond one's death is all but certain. Hence, by prolonging their own life, the agent would contribute nothing to the maintenance of the species.

2.1.4 'Enlightened pleasure-seekers' and accounts privileging mental self-maintenance

Finally, an account of virtue may be based on an account of function that privileges mental self-maintenance. Mental self-maintenance is closely related to positively valenced emotions, including the experience of desire-satisfaction, and the relative absence of negative emotions. Thus, it might seem that an account of function that privileges mental self-maintenance would be more promising. After all, it would resemble forms of virtue-ethics, such as that presented by Hurthouse (1999), according to which virtues contribute to eudaimonia, 'flourishing' or 'happiness'. We may call the agent whose faculty of practical reason functions optimally according to this account, an *enlightened pleasure-seeker*. Of course, the enlightened pleasure-seeker would not necessarily be a selfish egoist. After all, friendships likely contribute to self-maintenance. It is difficult, and perhaps impossible, to maintain friendships if you do not value the interests of your friends for their own sake. Similarly, a good standing in the community contributes to mental self-maintenance, both because the knowledge of one's good standing supplies one with a sense of ease and belonging, and because having a good standing in the community makes it easier to achieve many goals.

Nonetheless, such an organisational account of virtue based on an account of function which privileges the mental level could not give a successful account of ethical inquiry. Firstly, the choice of the privileged level is not scientifically justified. The mental level of self-maintenance is biologically

rather peripheral. With most species of living beings, it does not occur at all. Thus, the choice of the mental level of self-maintenance would not be driven by biological considerations, but purely by the goal of capturing certain pre-theoretical judgements about virtue. As such, however, it could provide little justification for these intuitions.

Furthermore, it cannot capture many of these intuitions anyway. Firstly, unlike the 'flourishing' described by Hursthouse, mental self-maintenance is completely internal. A mental system in Nozick's (1989) experience-machine simulating any sensory input a flourishing human in the real world would experience would be just as stable as the latter's. Thus, it seems that the enlightened pleasure-seeker would be disposed to accept the offer and enter an experience-machine that simulates her 'perfect life' if given the opportunity. Furthermore, experience seems to show that the wicked sometimes 'flourish like the green bay-tree'. Certain kinds of ruthlessness seem to be at least as good at leading to contentment as received virtue. In any case, empirical inquiry could easily show that this is the case. Even if it did, we would not be prepared to simply accept that such a life of ruthlessness is virtuous. Hursthouse (1999) avoids such objections by arguing that flourishing goes beyond subjective contentment. Vices like ruthlessness are intrinsically incompatible with flourishing. Such a move is not available to the mental-level-privileging organisational account.

2.1.5 Hypostatisation and complex weightings for levels of self-maintenance

One might try to argue that any such unacceptable implications could be avoided by a more complex account of function which weighs different levels of self-maintenance against each other in some stipulated way. For example, it may say that one year of added individual life-expectancy is equivalent to 10,000 years added to the expected survival of the lineage. Similar factors would be added for other levels of self-maintenance.

It is indeed plausible that if the conversion rates are calibrated just right, such a complex account avoids any unacceptable consequences. However, it would still fail to explain ethical inquiry because it would 'hypostatise' certain ways of being human (cf. Chapter Two). One way to describe the problem with the account that privileges the level of individual survival is that the way of life of the 'health freak' is privileged in ethical inquiry without there being any ethical reasons for it. A question which we can ask in ethical inquiry is: 'Is the best human way to live a life that maximises longevity?'

Such a question may lead to sustained inquiry. According to the longevity-privileging organisational account, this is based on a mistake or a confusion. Such inquiry is actually otiose, because given the organisational account in question, the answer is trivially 'yes'. Since this aspect of our practice is based on a mistake, we should reform the practice so that such inquiry could not occur. However, such a 'reformed' practice would not be recognisable as ethical inquiry anymore. The same argument applies, mutatis mutandis, to the way of life of the 'brood parent', the way of life of the enlightened pleasure-seeker and the ways of life recommended by accounts of function that supply a sophisticated, weighted method to agglomerate contributions to self-maintenance on different levels.

Thus, any such account would turn ethical inquiry into a mere special case of scientific inquiry. The task of ethical inquiry would merely be to solve technical problems of how best to realise a fixed goal, e.g. maximising longevity. Hence, it would fail to account for the autonomy of ethics.

2.2 Levels of maintenance are *not* commensurable

Thus, one might try instead to base an account of virtue on an account of function that does not provide a method of weighing contributions to self-maintenance. Levels of self-maintenance are incommensurable under such an account. Such accounts of function are plausible for reasons unrelated to ethics. Accounts which privilege one level of self-maintenance to the exclusion of all others are bound to miss some pre-theoretically functional traits. For example, an account that privileges the level of individual organisms above all else would exclude traits that contribute to procreation. An account that privileges the lineage would exclude traits that contribute to individual self-maintenance towards the end of life. The function of my heart is still to pump blood, even if I am lying on my deathbed and won't procreate anymore or do anything to contribute to the viability of my existing offspring. However, accounts that involve more than one level of self-maintenance will have to weigh them in some way that is arbitrary and not motivated by biological considerations.

However, under such an approach, an agent's ethical deliberation alone would determine which of character traits are functional (and thus virtuous). Thus, nearly every character trait would be functional if it is endorsed in ethical inquiry. It merely needs to make a contribution to self-maintenance on some level. However, this would make ethical inquiry impossible. Ethical inquiry would not be an effortful attempt to 'get things right' and to avoid 'mistakes'. Instead it would be mere arbitrary "picking" (cf. Ullmann-Margalit 2017). Insofar as an agent treats ethical

inquiry as anything but arbitrary picking and invests time and energy into getting it right, they would be confused or mistaken.

According to the organisational account of function, a trait T has the function to X if (a) it contributes to self-maintenance by X-ing, and (b) T's X-ing is regulated for by some other subsystems of the containing organism. Given this, if two incompatible traits T_1 and T_2 contribute to self-maintenance on different levels, then which of T_1 and T_2, if any, is functional depends on which one is regulated for.

2.2.1 Virtue-inquiry as a general regulatory subsystem

In Chapter Four, Section 3.2, I have argued that ethical inquiry is a general regulatory subsystem. In this subsection, I argue that virtue-inquiry in particular, is a general regulatory subsystem. As a part of ethical inquiry, virtue-inquiry is responsible for its status as a regulatory subsystem. In the remainder of this chapter, I will describe how the bare function of virtue-inquiry as a general regulatory subsystem determines the internal, autonomous logic of virtue-inquiry. Through virtue-inquiry's regulation of the rest of ethical inquiry, it also determines the logic of the latter as a whole.

As described in Section 1, virtue-inquiry is that subset of ethical inquiry which does not merely instantiate character traits, or dispositions to treat considerations as reasons. Rather, in it, the agent inquires into which considerations to treat as reasons, i.e. which character traits are virtues.

Virtue-inquiry is a general regulatory subsystem. Firstly, virtue-inquiry is a regulatory subsystem. If I reach the conclusion in virtue-inquiry to treat a certain consideration as a reason, and this conclusion affects my actions, then some character trait is regulated for. For example, if I reach the conclusion that I should be braver in the face of danger and less cowardly, then virtue-inquiry will regulate for traits that make me accept danger and against traits that make me avoid it. In cases in which the former traits are not strong enough to determine my behaviour, virtue-inquiry (or a subsystem controlled by virtue-inquiry) responds to this failure and determines my behaviour in similar ways as those traits would have. Thus, it fits the definition of regulation introduced in Chapter Four.

Virtue-inquiry contributes to self-maintenance in general rather than to self-maintenance on any particular level. Firstly, unlike other regulatory subsystems it can regulate in favour of any level of self-maintenance. I can 'force myself' to perform any action. Secondly, virtue-inquiry contributes to maintaining effective agency. It prevents me from being 'tossed around' by different occurrent desires and the expressions of different functional subsystems. Thus, I could not consistently achieve my conscious goals. Effective agency, i.e. the ability to achieve my conscious goals, however, strongly contributes to self-maintenance on all levels. I deliberately keep myself content, fed, sheltered and alive, help my kin and benefit strangers.

Besides virtue-inquiry, there are other regulatory subsystems which regulate character traits and which contribute to self-maintenance only on particular levels. Consider the two conflicting character traits, say an acquisitive tendency to attempt to possess as many resources as possible, and a generous disposition to share resources with one's friends. There are three possible cases regarding how such regulatory subsystems may interact with. (a) Neither of the character traits is regulated for by such a level-specific regulatory subsystem. (b) Both subsystems are regulated-for by regulatory subsystems which each contribute to self-maintenance on the same level as the character trait they regulate for. (c) One character trait is regulated for by such a level-specific regulatory subsystem, but the other is not. In case (a), it is obviously virtue-inquiry which determines which character trait is functional, as no other regulatory subsystems are involved. In (b), virtue-inquiry also determines which trait is functional. After all, the different levels of self-maintenance are incommensurable. Thus, which of the two apparent regulatory subsystems is functional is determined by the subsystems regulating them. There might be other, level-specific subsystems regulating them. However, in that case, the same problem of incommensurability arises. The only way to stop the regress is through virtue-inquiry. Finally, in case (c), virtue-inquiry similarly determines which trait is functional, given that it interacts with the character traits at all. If it regulates for the character trait that is already regulated for by the level-specific regulatory subsystem, then this character trait will obviously be functional.

There could not be another such general regulatory subsystem besides virtue-inquiry. If there is a system seemingly regulating character traits, but being in conflict with virtue-inquiry, then this system would not be functional. It would undermine effective agency, i.e. the ability to act

consistently and pursue aims over extended periods of time. Not being functional, it would not actually be a regulatory subsystem.

Thus, this second option avoids the worry of hypostatisation. The worry of hypostatisation arose because an agent can seemingly ask themselves, e.g. 'I know that my disposition to avoid all alcohol is functional, but still: why let that disposition determine my actions?' and inquire about whether to let that disposition determine their actions. However the fact that the agents inquires about the question and suspends judgement means that the first part of the compound question is not true. Their disposition to avoid all alcohol is not functional. Hence, this apparent scenario cannot occur. Whenever an agent reasons about whether to act in a certain way, there is indeed no empirical consideration which, according to the organisational account of virtue, should foreclose such reasoning.

2.2.2 The danger of trivialisation

However, given this, it seems to follow that the conclusions of such virtue-inquiry could almost never be wrong. If I reach the practical conclusion to act like a vegetarian, then vegetarianism is functional --- and thus virtuous --- for me. If I reach the conclusion not to be vegetarian, then meat-eating is virtuous for me, and so on for almost any practical question. Assuming that practical reasoning is well-functioning and itself virtuous if it contributes to virtuous actions and supports virtuous character traits, it follows that practical reasoning is nearly always well-functioning. The only exception would be cases in which the character traits regulated in favour of does not contribute to self-maintenance on any level. However, given the diversity of levels in the human case, such cases would be very rare. Thus, it seems that virtue-inquiry could not be guided by anything in reaching a conclusion. Sure, an inquirer may take themselves to be guided by, say, considerations of animal welfare in judging that vegetarianism is virtuous. However, they could just as well have let themselves be guided by the pleasure of eating meat. Whatever conclusion virtue-inquiry reaches, it will be virtuous. Thus, it would be misguided to put any effort into trying to 'get things right'.

Thus, there is an apparent dilemma. If the account of virtue is based on an account of function that makes different levels of self-maintenance commensurable, then it is vulnerable to the hypostatisation worry. If it makes them incommensurable, then it seems to imply an unacceptable

extreme form of relativism. Under neither version of the account is practical reasoning as an effortful attempt to reach a correct conclusion and to avoid error possible.

This dilemma cannot be avoided by a simple 'mixed' form according to which different levels of self-maintenance are partially but not fully commensurable. An account of function may determine an acceptable range of methods to weigh contributions to self-maintenance on different levels without determining a unique correct method. However, this would merely mean that one or the other horn of the dilemma applies from case to case. If two conflicting dispositions are such that some acceptable methods of weighing make one count as functional and others the other, then they will be incommensurable. Which one is functional would be determined by virtue-inquiry. Inquiry about the question would amount to arbitrary 'picking'. If, however, one conflicting character trait contributes less to aggregated self-maintenance than the other under all acceptable methods, then this character trait is privileged for reasons that lie outside ethical inquiry. It would be pointless to inquire whether it is ethically preferable. Hence, the account would be 'hypostatising'.

3 Ethical inquiry according to bio-constructivism

In this section, I resolve the dilemma introduced above. I describe how ethical inquiry proceeds according to the organisational account of function. I will argue that some degree of relativism is entailed by the commitments of Chapter Two, i.e. those to the naturalist response to the challenge from science and to avoiding hypostatisation. Given this, it is plausible that the truth of ethical judgements about an individual depends on their prior ethical judgements. Then, I will describe various constraints on properly functioning ethical inquiry. Insofar as ethical inquiry endorses judgements violating such constraints, it is not functional. Hence, it cannot act as a regulatory subsystem. Therefore, the character traits that it endorses are not actually functional and thus not virtuous. The judgements in question are not true. Hence, ethical inquiry, as an attempt to reach true judgements, proceeds by revising prior judgements which the inquirer starts out endorsing.

3.1 The relative inevitability of relativism

In the preceding section, I have argued that the organisational account faces a dilemma between hypostatisation and a trivialisation of ethical inquiry. One aspect of this trivialisation was a seemingly unconstrained relativism.

This follows from the commitment to the naturalistic strategy against that challenge from science, a similarly naturalistic acceptance of science's ability in principle to explain human behaviour and the close link between ethical discourse and action. As we have seen, in ethical discourse, no reasoning can intervene between the recognition that I have overall reason to perform a certain action, and performing the action. Of course, this does not mean that we cannot discern intermediate steps between the recognition and action, including steps which can fail to occur. For example, plausibly, in order to act on the recognition, an agent must first form an intention and then act on that intention. However, reasoning can neither intervene between the recognition and forming an intention nor between the intention and action. Thus, this requirement does not amount to a form of motivational internalism about reasons. For all I am assuming here, an agent may completely fail to be motivated by a reason they take themselves to have. However, if this is the case, this is not because they reasoned or deliberated to the practical conclusion not to act.

Given this, it follows that the fact that an agent has a reason to act in a certain way must involve them in some direct sense. Otherwise, an agent could always reason as follows: 'It is true that I have overall reason to φ. However, that is not really a fact about me. Hence, I shall not φ.' For example, according to a form of metaethical utilitarianism advocated by some Cornell-style naturalists (e.g. Brink 1984, Boyd 1988), an action's goodness is identical with the property of producing the greatest happiness of the greatest number. Cornell-style naturalists deny 'Moral Rationalism', i.e. the claim that agents always have (at least a defeasible) reason to do what's morally good. If, however, they identified an action's property of being morally good with there being a certain kind of reason for the action, then they would be vulnerable to the following objection. An agent could reason like this: 'Of course I know that I have a moral reason to be vegetarian, however this is really just a fact about the effect which being a vegetarian has on others, it does have no bearing on me. Thus, I shall not be a vegetarian'. Similarly, according to etiological functionalism, an agent has reason to act in a certain way if a species-typical human being would act that way, i.e. a human being realising the selected-for

functions of its faculty of practical reason. However, an agent could reason: 'Of course I know that I have reason to have children, but that is really just a fact about my ancestors that has no bearing on me. Thus, I shall not have children.'

The point is not that such reasoning is correct (though it might well be). Rather, the point is that it is intelligible. In contrast, in our form of ethical discourse, such reasoning is not possible. Hence, such accounts cannot explain the connection between ethical reasoning and action described in Chapter 1. They would hypostatise a certain standard of action, i.e. remove it from ethical critique even though we can, in fact, criticise it. Thus, a successful account of ethical discourse must show that its content has a bearing on the agent. I am not suggesting that a proposed account of rational normativity concerning the agent in this sense is sufficient for it to avoid the worry of hypostatisation. After all, there are many other reasons for me to dismiss a purported reason for action than that it doesn't concern me. Rather, it is necessary for it to avoid the worry.

It is, admittedly, somewhat vague what the relevant notion of *bearing* consists in. This is because it is itself an ethical question. However, it seems that the objection is always intelligible if the fact the consideration is not connected to local facts about the agent. One might respond to the claim that 'this has no bearing on me' by pointing connections between local facts about the speaker and the consideration in question. One such connection consists, of course, in the agent's motivational set: her preferences, wishes, desires, etc. One way to respond to the claim that a purported reason has no bearing on the agent is to point out how it will contribute to the satisfaction of the agent's preferences, etc. However, we may also point out how the purported reason relates to other local facts about the agent, such as their health. 'It will make you sick if you don't do it' is a possible response to 'it has no bearing on me'. It may be that such responses can be reduced to responses that refer to the agent's motivational set, but this is not obviously so. If I say 'this does have a bearing on you, it will make you sick if you don't do something about it', it is no valid counter to say: 'still, it has no bearing on me, I don't care about getting sick'. Seemingly, even if an agent doesn't care about their health, facts that affect their health still have a bearing on them.

Meta-ethical positions that do not adopt the naturalistic strategy in responding to the challenge from science may appeal to universal considerations, i.e. to facts that are local to all human agents. Such facts may concern, for example, the human 'life-form'/'species-essence' (e.g. Foot 2001; cf.

Thompson 2008), or a universal form of reason (e.g. Kant 2019 [1785], Korsgaard 1996). The life-form or the form of reason is local to every agent and has a bearing on every agent. However, such universal facts about human beings are excluded by the theory of evolution. The distinctions between species are gradual, including the distinction between human and non-human. In fact, with the possible exceptions of logical and mathematical truths and the laws of nature, modern science rejects any facts which are present and local at multiple locations. Logic, mathematics and the laws of nature, however, are unlikely candidates for providing practical reasons. Hence, the naturalistic strategy implies that facts about reasons are relative to agents: relative to what has bearing on them, and to what is local to them.

This argument is a more modest cousin of an argument from 'reasons-internalism' to relativism (cf. Mackie 1977, Williams 1981, 1985, Street 2012). According to that argument, it must be possible for an agent to be motivated by a reason for her to have that reason. Thus, if the contents of this glass being gin is a reason for me to drink them, then I must be capable of being motivated by the consideration that the contents of this glass are gin. Very plausibly, in this case, I can only be motivated in this way if I am already motivated to drink gin. Williams assumes that this is the case for every reason. Thus, what counts as a reason depends on an agent's motivational set, i.e. the collection of desires, wishes and similar motivational attitudes that they have. Furthermore, given a naturalistic, evolutionary understanding of human beings, we should not expect that there are any elements of such motivational sets that are truly universal among human beings. Thus, facts about reasons are relative to agents' motivational sets.

This argument can be attacked since it excludes the possibility that, if an agent comes to know a fact, she is motivated by the knowledge of that fact independently of any prior or independently existing motivational attitude (see e.g. McDowell 1998). It can further be attacked for assuming that for an agent to have a reason it must be the case that she could be motivated by the reason.

My argument for relativism is weaker in two ways. Firstly, it does not assume reasons-internalism. For all I am assuming, an agent may fail to be motivated by a consideration which counts as a reason for him. However, such an absence of motivation may not be due to reasoning. It may be due to, say, weakness of will or 'madness'. Secondly, I am not assuming that judgements of fact are never intrinsically motivating. However, all such motivating judgements must be about something that has

a bearing on the subject. Thus, facts about reasons are not necessarily relative to a subject's motivational set. Rather, they are relative to her more generally, i.e. to the things that have a bearing on her.

Thus, there is independent reason to think that the truth of ethical judgements is relative to the agent performing such an action. This removes one objection to the second horn of the dilemma between hypostatisation and trivialisation. It is true that if the organisational account of virtue is based on an account of function that leaves different levels of self-maintenance incommensurable, then whether a given character trait counts as virtuous at least in principle is relative to the agent whose trait it is. However, it turns out that we have independent reason to believe that this is the case. Thus, it is at least plausible that it depends on the ethical judgements the agent has endorsed which of their character traits are virtuous. One way for a consideration to have a bearing on an agent is for it to be a consideration which the agent would treat as a reason for action if they had a character trait which they believe they should have. If I say 'It is true that this person is suffering, but this has no bearing on me', you can felicitously respond 'but you agree that you should be more empathetic and responsive to the suffering of those around you, so it has a bearing on you'. Of course, for all that has been said so far, I may revise the latter judgement rather than the former to revise this inconsistency. Still, the judgement 'this consideration has no bearing on me' is incompatible with the judgement 'I should treat this consideration as a reason'. Beyond this initial plausibility, the claim that it depends on the ethical judgements the agent has endorsed which of their character traits are virtuous will be supported by the way in which it features in a way to explain various formal features of ethical inquiry.

3.2 Constraints on maximally functional virtue-inquiry

In the last sub-section, I have argued that the combined commitment of responding to the challenge from science using the naturalist strategy while avoiding hypostatisation entails that ethical truth is relative to agents. In particular, it is relative to the ethical judgements an agent accepts. Thus, one might think that 'ethical inquiry' must be trivialised. Any judgement one might make would be true. However, this is not the case. Rather, the dependence of ethical truth on ethical judgements is constrained. In order to act as a regulatory subsystem, ethical inquiry must itself be functional. It must contribute to self-maintenance in a consistent and systematic way. A character is functional not

if it is reinforced by ethical inquiry as it actually is. Instead, it is functional if it *would be* reinforced by ethical inquiry if it were functional.

Most of ethical inquiry is regulated by virtue-inquiry. Virtue-inquiry itself does not have a regulated function. There is no other subsystem which could regulate it. Yet, virtue-inquiry has a *bare function*. It contributes to self-maintenance in a consistent and systematic way. It does so by maintaining effective agency and reputation with other members of one's social group. In cases where it fails to do so, it is not malfunctioning. It is simply not functional. However, in cases where it fails to be functional, it does not count as regulating character traits.

Using the idiom of possible worlds, we can say that a trait is functional if it is reinforced by virtue-inquiry in the closest possible world where virtue-inquiry is *maximally functional*. This is the world where it makes the same kind of systematic contribution to self-maintenance that it actually makes, but it makes it as consistently as possible and to the greatest extent possible. Thus, constraints on ethical inquiry can be derived from the actual systematic contributions ethical inquiry makes to self-maintenance on all levels.

This model may strike some readers as needlessly convoluted. However, firstly, it is introspectively plausible. Sometimes, in ethical inquiry, we simply take some considerations to be reasons without thinking much about why we do so. Sometimes, we inquire about which considerations to treat as reasons. Furthermore, we could not simply identify virtues with the *bare* functions of character traits. Given the different levels of self-maintenance, almost any character trait will actualise some bare function. None of them contributes to self-maintenance *in general* the way virtue-inquiry does. Thus, almost any character trait would be virtuous. Vice versa, we could also not appeal to virtue-inquiry directly determining action. In this case, virtue-inquiry would have no ordinary subject matter and bio-constructivism could not respond to the challenge from science.

In the remainder of this section, I will argue for various such constraints on functional virtue-inquiry. I will first argue that there is a 'diachronic consistency constraint' on ethical inquiry. Other things being equal, an agent should continue to endorse the ethical judgments they have endorsed in the past. They should not revise their judgements arbitrarily. However, there are other constraints. These constraints correspond to functional ways to revise ethical judgements. Other

things being equal, in functional virtue-inquiry, ethical judgements are consistent. Character traits that are judged to be functional must be compatible with 'thriving', i.e. with basic good health and contentment. Realising the demands of ethical judgements must be feasible. Finally, ethical judgements must typically be at least potentially 'public', i.e. it should not be self-undermining for me to share this judgement with others. Each of these constraints correspond to 'moves' which we perform in ethical inquiry. Together, they constitute a distinctive *logic of ethical inquiry*.

4.2.1 Consistency constraints

In this section, I will discuss two types of 'consistency constraints'. These are diachronic consistency constraints and synchronic consistency constraints.

3.2.1.1 Diachronic consistency constraints

In order to be a general regulatory subsystem, ethical inquiry must contribute to self-maintenance in a way that is not specific to any particular level of self-maintenance. Otherwise, it would merely be one among many competing potential regulatory subsystems. It would not determine the function of the traits it regulated. The way that ethical inquiry contributes to self-maintenance is by producing effective agency. Effective agency must be consistent. Without the capacity to regulate their actions through ethical inquiry, the behaviour of human beings would be determined by different, competing traits. The complexity of human life is only possible because human beings have various different subsystems which may determine behaviour so as to benefit self-maintenance at various different levels. Thus, at one point, an 'altruistic' subsystem may dominate, which pushes me to give my resources to the most needy. At other times, a more selfish subsystem might dominate. In many cases, the result would be that my actions are self-undermining. I would succeed neither in benefitting my own survival or mental self-maintenance, nor would I succeed at benefitting the maintenance of a lineage or quasi-lineage of mine.

Through the regulating influence of ethical inquiry, such self-undermining can be avoided. Thus, ethical inquiry makes a general contribution to self-maintenance. In a particular situation, it contributes to self-maintenance on, say, the level of the individual organism without harming self-maintenance on other levels. In general, over time, it contributes to self-maintenance on all levels. However, this is only the case if the conclusions of ethical inquiry are consistent over time. If, at one point, I reach the conclusion that being altruistic is good, but at another that being selfish is

good, then my behaviour would be as self-undermining as it would be without ethical inquiry. Hence there are *diachronic consistency constraints* on functional virtue-inquiry. Other things being equal, I ought to continue to endorse ethical judgments which I have endorsed before. Thus, ethical inquiry takes the shape of revising and extending prior ethical judgements, rather than deducing ethical judgements 'from the ground up'. We all find ourselves always already endorsing some ethical judgements. These may be the conclusions of earlier inquiries, innate or the result of our upbringing.

Diachronic consistency constraints are in tension with ethical "conversions" (cf. McDowell 1995), but they do not make them impossible. An agent may undergo an episode that does not involve ethical inquiry, but in which their ethical judgements fundamentally change. As long as such episodes are rare in an agent's life, diachronic consistency constraints will count in favour of relying on the *later*, post-conversion judgements after the conversion. Doing so may lead to some self-undermining behaviour. However, attempting to return to the older, pre-conversion judgements through ethical inquiry would likely be only partially successful and lead to more self-undermining. Hence, at least retrospectively, such conversion-episodes will appear to be neither rational nor irrational as far as the logic of ethical inquiry is concerned. In contrast, if conversion-episodes are too frequent in a person's life, they may begin to seriously undermine their effective agency. Hence, maximally functional virtue-inquiry *would* attempt to revise them. They would be irrational.

3.2.1.2 Synchronic consistency constraints

However, diachronic consistency constraints are not the only constraints on maximally functional virtue-inquiry. These other constraints explain how ethical inquiry is possible. Without them, apparent 'virtue-inquiry' would only consist of re-affirming one's prior ethical judgements.

One such set of constraints are ***synchronic*** *consistency constraints*. Just as arbitrarily overturning prior ethical judgements would undermine my ability to consistently pursue goals and advance self-maintenance that way, so does simultaneously endorsing inconsistent, contradictory or otherwise mutually undermining ethical judgements. By inconsistent ethical judgements, I mean judgements which describe the virtuous agent in incompatible ways. If I judge both that one ought to be charitable towards the most needy and a rugged individualist who only looks out for themselves above all, then these judgements are inconsistent. Other things being equal, if I simultaneously endorse incompatible ethical judgements, my conduct will be self-undermining. Hence such ethical

inquiry would not be functional. Thus, in ethical inquiry we revise our prior ethical judgements so as to make them more consistent.

In fact, such constraints do not just count in favour of mere non-contradictoriness among ethical judgements. Rather, an agent is more effective at pursuing their goals (any goals) if their actions positively reinforce each other and if they pursue similar, related or complementary goals in many of their actions. Thus, functional ethical inquiry not only eliminates inconsistency, but it promotes *coherence* among ethical judgements. Such coherence may be achieved by revising one's ethical judgements such that more of them can be derived from central principles. Thus, we follow the methodology described by Hursthouse with reference to Neurath's boat.

There are (at least) two strategies of revising judgements in this way. We may, firstly, attempt to provide descriptions of *desirable outcomes* and then judge individual actions and character traits based on the extent to which they promote these outcomes. Secondly, we may try to identify general rules one ought to follow and/or types of actions which are always blameworthy or always praiseworthy. Hence, both consequentialist and deontological reasoning have a place in ethical inquiry. However, all such strategies must be weighed against other constraints. This does not mean that deontological or consequentialist ethical theories are necessarily incorrect. They *may* be correct, if they can be shown to be correct in virtue-inquiry respecting all relevant constraints.

3.2.2 Empirical Constraints

So far, the constraints that I have considered only concern the relations of ethical judgements among each other. In this subsection, I will introduce two constraints that are based on empirical considerations instead. These are 'feasibility constraints' and 'thriving constraints'. The constraints are empirical insofar as through them, functional ethical inquiry is constrained by facts unrelated to what ethical judgements the inquirer endorses or has endorsed. Hence, in making one's best effort to reach functional (and thus true) conclusions in ethical inquiry, non-ethical empirical judgements will be relevant to ethical inquiry.

Through such empirical constraints, bio-constructivism can explain the cases of so-called "Ideally Coherent Eccentrics (ICEs)" (Street 2009). These are agents whose attitudes are perfectly consistent,

but nonetheless so bizarre that it seems unacceptable to say that they are completely reasonable, i.e. that they have no reason to act differently from the way they do.

The relevant empirical judgments concerning patterns in human behaviour are discovered in a form of discourse that is so closely related to ethical inquiry in a narrow sense that it can only theoretically be disentangled. Scientific inquiry into human behaviour is of course relevant to this inquiry. However, human behaviour is so complex that it is inevitable that a vast majority of the significant patterns will not be discovered using scientific inquiry. Doing so would require collecting vast amounts of data and the capacity to analyse that data. This includes patterns that are relevant to self-maintenance, such as which character traits are likely going to lead to contentment and self-maintenance on the individual level for which agents, and which are likely to sustain lasting cooperation.

Of course, human ethology, psychology, sociology, economics, etc. have produced a lot of scientific knowledge regarding such patterns. Such knowledge is indeed relevant to ethical inquiry. However, such knowledge is much more limited than the knowledge an agent can gain through other ways. We are almost constantly interacting with other human beings and thus gather an immense amount of evidence concerning patterns in their behaviour. However, this evidence cannot be collected and documented in a way that would be eligible in scientific discourse. Furthermore, we have specialised skills to interpret and organise such evidence concerning human behaviour. These likely depend on specialised adaptations. This is evidenced, for example, by the fact how useful it is for us to portray processes and patterns that do not involve human agents as if they did involve human agents. Likely, some of these skills depend on selected-for capacities. This involves representations in the form of narratives. However, given that we cannot (yet?) represent such interpretations formally, it is not eligible in scientific discourse.

Thus, there is an 'ascientific' way of inquiring into what is feasible for human beings, i.e. a way that is different from science, but not incompatible with science. This form of inquiry involved imagination, given that my capacity for imagination is informed by my experiences of human life. Thus, such inquiry likely involves, for example, fictional narratives. It may also involve so-called "experiments in living" (Mill 2014 [1859], Anderson 1991). Ethical inquirers may decide to live in a different way than has been done before. For example, they may decide to raise their children

without the use of corporal punishment and other forms of physical violence. Thus, they discover new facts about patterns in human behaviour. For example, they may discover that children who have been raised without physical violence do not typically grow up wicked, undisciplined and self-indulgent, as would have been supposed before.

3.2.2.1 Feasibility constraints

The first empirical set of constraints are *feasibility constraints*. It must actually be possible for human beings like the inquirer to act in the way that is recommended by their ethical judgements, both taking the judgements individually and considering them jointly. For example, suppose I judge that it would be most virtuous for me to become a hedge fund manager, work 70 hours per week and donate 95 percent of my income to malaria prevention. However, it would actually be impossible for me to maintain the energy and motivation that is required for such a lifestyle. In this case my ethical inquiry would not be functional, and hence the conclusions would be false. Such feasibility constraints apply both to individual human traits and to social arrangements. If I judge that it is virtuous to promote a certain social way of living, yet this way of life is not actually feasible for human beings, then my ethical inquiry is not functional. The consideration that a way of life is not feasible counts as a reason against judging that this way of life is uniquely virtuous.

Ethical inquiry is constrained by what is feasible for the agent whose actions and character traits are under consideration. That is, in the ordinary cases of ethical inquiry we are considering so far, the inquirer themselves. However, we are likely to deceive ourselves when we consider what is feasible for ourselves in particular. Consequently, ethical inquiry is typically more likely to be functional if I revise my ethical judgements to be compatible with what I judge to be feasible for human beings broadly like me, rather than what I think is feasible for me in particular.

3.2.2.2 Thrivingness constraints

Another set of empirical constraints are *thrivingness constraints*. By *thriving* or *thrivingness*, I mean a combination of health, on the one hand, and subjective 'happiness' or contentment on the other hand. That a trait precludes an agent from thriving counts as a reason that it is not virtuous. Very strong reasons related to other constraints are needed for a conclusion that violates thrivingness constraints to be endorsed in maximally functional virtue-inquiry.

Thriving is closely related to 'flourishing' or 'eudaimonia', as it has been described by virtue ethicists such as Rosalind Hursthouse (1999). In particular, it corresponds to an empirical component in the notion of flourishing. Hursthouse (1999: 186) describes the "smile factor" as one criterion by which we may judge whether an agent is "flourishing". Williams (1995: 46) similarly mentions an "ethological standard of the bright eye and the gleaming coat" which is supposed to indicate "flourishing". Hursthouse argues that acting virtuously does not typically make you worse off by the smile factor, though it might be tied with alternatives. Williams is more pessimistic. He suggests that being wicked often leads to more 'flourishing' than the alternative by such an empirical criterion.

'Health' closely relates to self-maintenance on the level of the individual. Leaving aside the special categories of mental health and reproductive health, an individual who is successful at staying alive is typically healthy, and a healthy individual is typically able to keep themselves alive except for surprising outside shocks. 'Contentment' or 'subjective happiness' is closely related to self-maintenance on the mental level. This is because positively valenced emotions have the (selected- and regulated-for) role of reinforcing an existing mental organisation. If I am feeling good, that means that things are going well, and not much 'needs' changed in the way I behave and in the way my behaviour is determined. The gloss 'contentment' further stresses that 'thriving' excludes states of extreme euphoria which may precipitate mental breakdowns rather than contribute mental self-maintenance. We would not ordinarily describe such people as 'contented', or as 'flourishing'.

Self-maintenance on the mental level and on the level of the individual organism have a special role. They are conditions for most contributions to self-maintenance on other levels, as especially insofar as these involve actions. An agent who does not live cannot reproduce. They also cannot benefit their lineages or quasi-lineages. Similarly, an agent who suffers mental breakdowns or depressive despondency is unable to act in ways that contribute to self-maintenance on any level. Hence, any trait that interferes with self-maintenance on these two levels is likely to interfere indirectly with self-maintenance on many other levels. Thus, any *direct* contribution that such a trait makes to self-maintenance on such other levels must be very strong for such a trait to be functional. Otherwise, the positive direct contribution will be outweighed by the many negative effects on self-maintenance on the *same* level which the trait will have *indirectly* through interfering with self-maintenance on the levels of thriving. Without making an overall positive contribution to

self-maintenance on *any* level, the trait in question cannot be functional. Hence, it cannot be 'regulated for' in the technical sense of the organisational account of function. An apparent regulatory subsystem reinforcing the constraints it exerts would not be functional. Hence, insofar as ethical inquiry endorses traits that interfere with thriving, it is likely not to be functional.

Some exceptions concern cases in which the contribution to the self-maintenance of the lineage, quasi-lineage, etc. is so strong that it outweighs the indirect negative effect. In such a case it is possible that successful ethical inquiry would regulate in favour of the trait that interferes with thriving. Thus, it may be virtuous to sacrifice your own life to save others.

Thus, there is another 'move' in ethical inquiry. If a trait heavily interferes with thriving, this counts as a reason to judge that it is not virtuous. The only way to preclude this inference is by showing that it provides a very strong benefit to others, i.e. to the lineage or the quasi-lineages.

3.2.2.3 Ideally Coherent Eccentrics

The term "Ideally Coherent Eccentric" (ICE) was coined by Street (2009) for characters featuring in thought experiments intended to challenge relativism, i.e. views according to which the reasons an agent has are relative to the motivational attitudes that they have. These characters are stipulated to be perfectly coherent in their conative states. Yet, they have preferences and/or perform actions which are so outrageous that people will find them unreasonable or ethically unacceptable. However, since ICEs are coherent, agent-relativists are committed to judging that they do not have reason to act differently from how they act. Given that bio-constructivism places a lot of importance on consistency constraints, critics might worry that it might similarly entail that ICEs are virtuous and reasonable, i.e. that they do not have reason to act differently than they do. In this section, I will consider two standard examples discussed by Street, namely the "Ideally Coherent Anorectic", due to Gibbard (2003) and the "Future Tuesday Indifferent", introduced by Parfitt (1984: 124, 2011).

3.2.2.3.1 The Ideally Coherent Anorexic

According to Gibbard (2003), the "Ideally Coherent Anorexic":

> "accepts norms that prescribe death by starvation, if the alternative is a figure plump enough to sustain life," and she is stipulated to be entirely coherent in doing so, such that she is

> making no mistakes whatsoever about the non-normative facts and "the norms she accepts tell her to starve, and the higher order norms she accepts tell her to accept those lower order norms" (summarised by Street 2009: 279).

As Gibbard insists, even if the anorexic thinks she is "rational", i.e. act as she has reason to, "the point here is, the rest of us do not. We think it irrational to starve oneself to death for the sake of a trim figure". In supporting 'our' claim that the anorexic is "irrational" we may point, for example, how she is killing and destroying herself, how she is not helping anyone, and how she is gaining nothing by starving herself --- not even fame, admiration or glory. These moves correspond to thriving and feasibility constraints. Firstly, by starving herself she directly and immediately undermines her physical self-maintenance. She also undermines her mental self-maintenance, since any mental self-maintenance depends on her continued bodily existence. In this, her behaviour contrasts with other unhealthy behaviours. Regularly drinking wine with friends may shorten my life expectancy. However, even with the habit, I will likely live long enough for it to contribute to my mental self-maintenance, and e.g., prevent mental breakdowns. In contrast, the anorexic will bring about her own death within years or months. Furthermore, the anorexic starving herself does not significantly contribute to the self-maintenance of any others. In fact, given the benefit she could have provided to others if she lived longer, she harms self-maintenance on the level of lineages and quasi-lineages as well. Thus, despite her coherence, the ideally coherent anorexic is almost certainly irrational. The considerations advanced to support this claim in ethical discourse correspond to facts which explain her irrationality according to the organisational account of virtue.

Street, in contrast, argues that the ideally coherent anorexic, as described by Gibbard, is rational. She supports this, firstly by arguing that no actual human being could ever be an ideally coherent anorexic. She argues that "[a]ny real-life woman, for example, presumably cares about the lives of her family and friends; she has a career or another broad plan of life that she hopes to pursue; she may be interested in travel or science or how the Braves are going to do next season; and so on." She would also endorse some higher-order norms endorsing these interests etc. She argues that actual people suffering from anorexia provide no counterexample to this claim for two reasons. Firstly, the real-life sufferers do not actually desire a figure incapable of sustaining life. Rather, they desire a "supermodel" figure, which is capable of sustaining life (as evidenced by supermodels). Such a figure, according to Street, however, is not attainable to many women. Thus, they fall into

desperately starving themselves. Secondly, they do not desire such a figure as an end in itself. Rather, they desire it as a means to many social goods.

Street's argument here is not convincing. Firstly, she goes on to compare the ideally coherent anorexic to "a person who knowingly accepts a tremendous risk of dying in order to have a shot at reaching the summit of K2". However, while they are rare, people like that exist. Why could there not be an agent for whom having an emaciated figure is as important as conquering K2 is to the climber, despite their desire to know "how the Braves are going to do next season"?

Secondly, the close association between anorexia nervosa and exposure to the celebration of "supermodel" figures has been questioned (cf. Corrington 1986, Arnaud et al. 2024). Anorexia nervosa may be more closely related to a need for *control,* especially in individuals who otherwise experience few ways to control their lives. As such, the desire of a person suffering from anorexia to control and conquer nature in the form of her own body may not be so dissimilar from the desire of an obsessive mountaineer to conquer nature in the form of wild peaks. Realising the desire for an emaciated body may really be an end in itself for some people, as it expresses a unique form of control. Thus, contra Street, real-life anorexics may be consistent. For Street, this would mean that they are not unreasonable. In contrast, according to bio-constructivism, they are unreasonable even if they are consistent, since their reasoning still violates the distinctive logic of ethical inquiry.

3.2.2.3.2 Future Tuesday Indifference

Secondly, Street considers the Future Tuesday Indifferent (FTI), a case due to Parfit:

> "Consider... an imagined man who has an attitude that we can call Future Tuesday Indifference. This man cares about his own future pleasures or pains, except when they will come on any future Tuesday. This strange attitude does not depend on ignorance or false beliefs. Pain on Tuesdays, this man knows, would be just as painful, and just as much his pain, and Tuesdays are just like other days of the week. Even so, given the choice, this man would now prefer agony on any future Tuesday to slight pain on any other future day."

Parfit argues:

> "That some ordeal would be much more painful is a strong reason not to prefer it. That this ordeal would be on a future Tuesday is no reason to prefer it. So this man's preferences are strongly contrary to reason, and irrational [even though he is perfectly coherent]" (via Street 2009: 281).

Together with some plausible assumptions about human beings, and any other organisms that evolved by natural selection, thrivingness constraints imply that FTI is unreasonable. This is because FTI will undergo many otherwise avoidable, extremely painful episodes. Pain evolved as a signal of damage to physical integrity and the ability to self-maintain. Therefore, pain will typically act as a signal to change one's conduct, to do something differently. 'Doing something differently' may simply mean removing one's hand from a hot stove. However, in case of strong, enduring or repeated pain, the pain will typically call for some fundamental changes of behaviour. If it endures longer, it can lead to a complete reorganization of one's mental system. If this re-organisation fails, it may be experienced as trauma. Thus, pain, especially frequent and intense pain is directly harmful to mental self-maintenance. In cases such as that of FTI, the frequent mental reorganisations would also likely impact FTI's ability to maintain themselves on other levels. For example, he might struggle to look after himself and supply himself with shelter and food, thus damaging his ability to self-maintain. In ordinary ethical discourse, we would likely not appeal to such long-term consequences in order to explain why FTI is irrational. We would simply say something like: 'But it will hurt!' However, the long-term consequences allow the organisational account of virtue to vindicate the role appeals to pain play in ethical discourse.

Thus, bio-constructivism captures one aspect of the apparent unreasonableness of FTI: their disregard for the 'intrinsic badness' of pain. It does not, it seems, capture another aspect of this unreasonableness. This FTI's *arbitrariness*. It is entirely groundless for them to be indifferent to pains on future Tuesdays rather than on any other day. It is true that bio-constructivism allows a fair degree of 'arbitrariness'. Ethical inquiry choosing between different levels of self-maintenance may often only be constrained by past judgements. However, FTI's arbitrariness is nonetheless relevant. Thrivingness constraints can be overruled if a trait that interferes with thriving --- such as Future Tuesday Indifference --- contributes very strongly to self-maintenance on some level. The arbitrariness shows that this is not the case.

In her discussion of FTI, Street distinguishes two versions of the thought experiment. In "Future Tuesday Indifference with Present Tuesday Horror (FTI-PTH)", FTI-PTH is indifferent to pain on Tuesdays when he anticipates it. Come Tuesday, however, he is horrified at what his past self has gotten him into. If he experiences serious pain on Tuesday, he is agonised by it. In contrast the "Consistent Tuesday Indifferen[t] (CTI)", is unfazed by the pain. They experience it, but they are detached from it in the way some meditators may be detached from pain and other sensations they have during meditation. Street argues that CTI is rational. This assessment is plausible. CTI's being unfazed by the pain may indicate that it does not have the deep and lasting effects on his mental organisation it would otherwise have.

Street suggests that FTI-PTH is unreasonable for the same reasons as an immoral agent. He relates to his 'Tuesday self' more as one would relate to a different agent. For example, non-Tuesday FTI-PTH and Tuesday FTI-PTH frequently undermine each other's projects. Given this, non-Tuesday FTI-PTH is treating Tuesday FTI-PTH in a reprehensible way. In another discussion, Street (2012) argues that ideally coherent agents who violate our deepest moral convictions can indeed be rational. However, we have strong reasons that no human beings are like this. However, it is dubious whether non-Tuesday FTI-PTH and Tuesday FTI-PTH are really separate persons in a morally significant way. For example, they share the same body, and while some of their projects are different, others may be the same. They also share the same memories. Thus, the organisational account's way of explaining the irrationality of FTI-PTH is superior.

3.2.4 Shareability constraints and impartiality

The cases of "Ideally Coherent Eccentrics" considered above concern cases in which agents are unreasonable because they lack self-regarding virtues. Yet, perhaps more troubling cases concern what Street (2009) calls the "Ideally Coherent Caligula". (In her 2009, she merely mentions that she will discuss such a character in a following publication; while she does not use the term again, this publication seems to be Street 2012). This is an agent who intuitively lacks other-regarding virtues. They desire cruelty for its own sake and rejoice in harming others. Yet, they do so without being inconsistent.

In this section, I will introduce a set of constraints that help explain how such a character is irrational. I call these *shareability constraints*. So far, I have made the simplifying assumption that in ethical inquiry, there is a singular inquirer. In reality, ethical inquiry is typically shared with others. I present ethical arguments (implicitly based on the constraints considered in this section) to others, and they evaluate these arguments. Similarly, I may evaluate arguments presented by them. Such interactions may influence my behaviour. For example, I may get convinced of ethical veganism and regulate my practical deliberation so as to consider my actions impact on the well-being of animals. I may also adopt such a position simply because I observe others around me adopting it.

Ethical inquiry being public in this way has various advantages regarding self-maintenance. Firstly, ethical inquiry can be costly. It takes time, attention and energy to conduct ethical inquiry. Offloading it into socially distributed systems saves such resources and allows them to be devoted to other ends that contribute to self-maintenance. Mercier and Sperber (2017) argue that in all forms of inquiry, human beings are generally good at generating arguments for conclusions they favour, and at impartially evaluating others' arguments. Yet we are bad at evaluating arguments they generated. This is because these are the *functions* of our capacity for explicit reasoning (see also Henriques 2003). Secondly, such socially distributed ethical inquiry allows others to contribute their empirical knowledge to ethical inquiry.

Furthermore, given that ethical inquiry regulates our actions, it makes my actions more predictable to others, and it makes their actions more predictable to me. Thus, it enables cooperation. Finally, if ethical inquiry is public in this way, others are likely to support the ways through which ethical inquiry regulates action. For example, they may remind me of the conclusions that I have reached in ethical inquiry when I am likely to act contrary to them. They may also punish or sanction me if I act contrary to the judgements we share.

Ethical inquiry can only be public and shareable in this way if the conclusions which I reach are conclusions which I may share with others without undermining my ability to reach my goals. For example, if I reach the conclusion that it is virtuous to deceive others for my own gain, I could not share that conclusion without undermining my ability to reach my goals. Doing so would warn people of my intention to deceive them.

The same is at least sometimes true of intentions to use violence against others. If it was widely known that these were my intentions, others would avoid me or prepare to defend themselves against me. At other times, such violent intentions may be perfectly public. A member of a protection racket may *benefit* from making his intentions public, as it intimidates their victims. However, they still cannot share their ethical inquiry in the same sense with their victims. This is because I could not expect them to engage in genuine ethical debate with them. Rather than evaluating my arguments, they would simply try to convince me not to use violence against them. Doing so is both what we would expect from them, and what would contribute to their self-maintenance.

One might worry that it is *always* in the interests of others to use ethical debate in order to manipulate the primary inquirer rather than evaluate their arguments. However, firstly, evaluating others' ethical arguments has advantages for oneself. If the arguments are true, one can accept the conclusion oneself and has gained ethical knowledge. Secondly, attempting to manipulate others carries the risk of sanctions if found out. I might be condemned as a liar or a hypocrite.

Furthermore, such an intention will itself violate shareability constraints. I would have to keep it secret from others. Otherwise, they would not rely on me anymore in ethical debate. Thus, at least with regard to this intention; I could not rely on others in ethical debate. Finally, engaging in genuine ethical debate with others will typically make their behaviour more predictable than attempting to manipulate them. If I manipulate them, they may discover that they are being manipulated and/or succeed in ethical inquiry without external support and change their behaviour. In contrast, the likelihood that they will change their behaviour if they are already engaging in successful ethical inquiry is much smaller. Other things being equal, it is advantageous for me to be around others whose behaviour I may predict. Thus, it is only with especially extreme intentions of harm towards others that I could not generally expect them to engage in genuine ethical debate with me.

Thus, other things being equal, ethical inquiry is not functional if it leads us to endorse conclusions which must be kept secret from others in order for the agent to achieve their goals, or where it would be clearly functional for others to attempt manipulating into abandoning them rather than engaging in genuine ethical debate. This is, of course, a *much* weaker version of Kant's famous imperative only to act according to maxims which every rational being can rationally endorse. I

suspect that Kant's project derives part of his plausibility from this constraint. However, unlike Kant, I do not argue that all of morality can be derived from this constraint alone. Consequently, I do not believe that this constraint is absolute. Sometimes, other constraints might outweigh. Thus, being prepared to steal in order to feed one's starving family may easily be virtuous, even if it requires me to keep my judgement secret from those I steal from.

Shareability constraints explain several moves in virtue-inquiry that tend to reinforce impartiality. The first move can be encapsulated in the question: *what if everyone did this?* Suppose an ethical conclusion is such that if everyone adopted it, the consequences would harm my own interests and goals. Pursuing it may require keeping it secret, since otherwise it might inspire others to adopt the same conclusion, leading to the undesirable consequences. Thus, if a conclusion being generally adopted would have such consequences, this counts against adopting it. Furthermore, even if I do not keep my conclusion secret, others will be more likely to use ethical debate to manipulate me, and doing so will be more likely to be functional. After all, it is in their interests to prevent the judgement from spreading. Levine et al. (2020) show that this question is indeed a common move in ethical inquiry.

Secondly, shareability constraints support the principle *never to make an exception for yourself*. If you judge that others ought to behave in a certain way, then you also ought to judge you ought to behave in that way, and vice versa. This is so unless you can point at some morally significant difference between yourself and others. This principle is related but different from the move of asking *what if everyone did this?* After all, it may be undesirable for me if everyone acted a certain way, but I can judge that they would not be vicious for acting that way.

If I make an exception for myself (or those near and dear to me) in this way, then I cannot benefit from ethical debate. This is, firstly, because it would be in my interest to keep such judgements secret. If others knew about me making judgements, this would make cooperation with them difficult or impossible. They would reasonably suspect that I am exploiting them in the exchange, i.e. treating them in a way where it is to their benefit not to cooperate with me. Thus, in order to maintain my ability to benefit from cooperation, I would have to keep those judgments secret. Furthermore, even if I do not keep my judgments secret, others would be less likely to engage in genuine ethical debate with me. This is because even if they did not find any flaws with my

arguments, they couldn't accept them for themselves to guide their behaviour. The conclusion would not concern them.

This principle of never making exceptions for oneself applies both in a holistic and in an atomistic way. In a holistic way, it ensures that my judgements of others constrain my judgements of myself and vice versa. As an atomistic way, it excludes judgements which on their own require making an exception for oneself. These are judgements in which the behaviour prescribed itself depends on others' not adopting the same judgement. Take the judgement that it belongs to a virtue of competitiveness to kill one's rivals in order to attain and enjoy a promotion (cf. Korsgaard 1985). If others were to act the same way, I could not enjoy my promotion, as I'd be the next target for assassination.

This constraint is more narrow than the question: *what if everyone did this?* That's because it may have undesirable consequences for everyone to adopt some ethical judgement without that judgement being self-undermining. It may be undesirable if nobody were to show up for their friend's birthday if they can watch Netflix on their own instead. After all, people would be lonely on their birthdays. Yet, it would not make watching Netflix at home impossible or significantly less desirable. Thus, there are additional, stronger grounds for rejecting such self-undermining conclusions in virtue-inquiry than there are for other conclusions that would have undesirable consequences if everyone adopted them.

Finally, shareability constraints support reasoning according to the *Golden Rule*: 'Do unto others as you would have them do unto you' or, in its negative version, 'Do not do unto others as you would not have them do unto you'. Assuming that others have broadly the same interests as myself, if I make a judgement that allows me to act toward others in a way I would strongly dislike for them to act towards me, then I should keep this judgement secret. Otherwise, I may expect them to manipulate me rather than engaging in genuine ethical debate. Furthermore, if I do share the judgments, others would be less likely to engage in genuine ethical debate with me rather than to manipulate me into abandoning the judgement. Hence, it is a move of ethical inquiry to ask 'would I like to be treated that way?' when considering a possible course of action concerning another agent. Versions of the Golden Rule are known in various societies and ethical/religious traditions around the world. This includes Daoism, Confucianism, Islam, Judaism, Christianity, Buddhism, Jainism,

Mandaeism, ancient Egyptian wisdom literature as well as the oral Yoruba and Igbo wisdom traditions (see Gensler 2013, ch. 3).

Thus shareability constraints provide the foundation of various principles of impartiality that are embedded in the logic of virtue-inquiry. These principles of impartiality do not prescribe complete impartiality in my actions. It most likely belongs to the virtue of prudence, sometimes to be partial to oneself and to the virtue of loyalty to be partial to one's friends. However, if I judge that I ought to be partial in this way, I cannot consistently with the logic of ethical inquiry judge that others ought not be partial in these ways.

An important question of such principles of impartiality is the question of the scope of impartiality. Who are the 'others' in the Golden Rule? Who is 'everyone' in the question *what if everyone did this*? What property would justify an exception that applies to me or my near and dear? In some societies, such as Ancient Greek societies, such moves have been recognised, but they were only applied to other free Greeks. It is an open exegetical question to what extent the injunction to 'love thy neighbour as thyself' as it occurs in the Hebrew Bible is meant to extend beyond fellow Israelites (see Gensler 2013). Another, related question concerns which aspects of the other's situation ought to be taken into consideration. A contemporary liberal would likely say that their interests and their past actions ought to be taken into consideration. Thus, a judge may punish a criminal to jail time even though they themselves would not like to be thrown into jail. They have not committed a crime. However, their sex, gender, caste, race or socio-economic status ought not to matter.

The source of shareability constraints create some pressure regarding the scope of impartiality. Typically, the people involving whom in ethical inquiry has the greatest benefit for self-maintenance are those who are most directly involved in my actions. They are most likely to have empirical knowledge that is relevant to my inquiry. As I am interacting with them, successful cooperation with them has the greatest potential benefit. They are also best positioned to support me in acting on the conclusions of my ethical inquiry. Thus, other things being equal, I should at least be able to share my ethical inquiry with those involved in my actions. As we will see in more detail in Chapter Seven, this accounts especially against justifications of enslavement, in which the enslaved are excluded from the scope of impartiality. However, the exact scope can only be determined in ethical inquiry shaped by all of the constraints, including consistency constraints.

3.2.5 What if the constraints conflict?

One may wonder what happens if the different constraints conflict. However, given that most of the constraints simply constitute *minimal* thresholds rather than constraints of maximisation, there are only few opportunities for constraints to conflict. Of course, the diachronic consistency constraint regularly conflicts with the other constraints. In order to satisfy the latter, an agent may need to revise some of their judgements. In these cases, in principle, the latter constraints take precedence. In order to actualise its function at all, ethical inquiry must be capable of revising judgements. Hence, if the diachronic consistency constraint were to take precedence, it could not actualise its function. Thus, the diachronic consistency constraint is mostly only relevant if there are no other relevant constraints. It prevents arbitrary revisions of ethical judgements. It also prescribes a preference for more conservative revisions in satisfying the other constraints.

That being said, the virtue of caution in epistemic inquiry may prescribe adopting new ethical judgements only piecemeal, even if one suspects that many of one's prior ethical judgements are incorrect. This virtue matters as to how one practically ought to proceed in revising one's judgements. However, it does not affect which maximally functional configuration of judgement's one would eventually reach. Hence, it does not play a role in determining ethical truth.

Feasibility, thrivingness and shareability constraints are satisficing constraints. They do not prescribe maximising any metric, but rather meeting a sufficient standard. Hence, typically, it will be possible to revise your ethical judgements in such a way that you can publicly share your judgements, that you can be minimally healthy and content while living according to them and that it is feasible for a society to operate according to them. In some cases, this may be impossible. For example, I may be required to adopt a policy of dishonesty in order to remain satisfactorily healthy. In such a case, one must decide in light of one's other ethical judgements whether to revise one's judgements so as to satisfy one constraint or the other. Thus, the synchronic consistency constraint resolves the conflict.

4. Converging and judging others

In the foregoing sections, I have focussed on *self-directed* ethical inquiry: inquiry into how *I* should act and what *I* should be like. However, we also frequently make judgements about, and inquire into, how others should act. In such judgements, we often rely on the same ethical conclusions as we do

to regulate our own actions. In this section, I argue how such inquiries are justified concerning fellow community members and members of communities we closely interact with. In Chapter Six, I will focus on evaluations of members of distant communities.

In Section 3.2.4, I have already argued that, typically, ethical inquiry is public. This allows me to draw on the support and knowledge of others in order to reach conclusions in ethical inquiry, and in order to act on those conclusions. In addition, it is actually beneficial for self-maintenance for fellow community members to regulate their behaviour based on shared or identical ethical judgements. Such shared judgements make it easier for members of the community to predict each other's behaviour, and thus to cooperate. It also makes it less likely that members of the community would undermine other's efforts, for example in helping outsiders or in contributing to the preservation of the natural environment.

Given this, it is functional for ethical inquiry not only to treat one's own judgements as 'priors' that should be accommodated. Rather, functional ethical inquiry treats the judgements of *other* community members with whom I regularly interact in a similar way. If there is an ethical disagreement between me and a fellow community member, the diachronic consistency constraint does not allow me to simply dismiss their judgement since my judgement is the one I find myself with. Rather, I ought to attempt to resolve the disagreement through ethical inquiry based on the other constraints. The same is true for any party in such an agreement. This does not mean that agreement will be eventually reached. For example, I may maintain my judgement that everybody else in my community is treating animals inappropriately if I find this judgement to be well-supported in ethical inquiry, and I am not convinced by any of my fellow community member's arguments.

Hence, ethical inquiry for different members of such a community relies on the same pool of 'prior' ethical judgements. It is largely irrelevant for a member of the community engaging in ethical inquiry whether an ethical judgement has been endorsed earlier by *themselves* in ethical inquiry or by another member of their community. Thus, insofar as for each of these members, functional ethical inquiry has the same basis of prior judgements, insofar as their ethical inquiry is functional, their judgements would converge. Hence, it is appropriate for members of the community to rely on their own ethical

inquiry to judge other members of the community. Of course, in turn, they ought to sharpen their own ethical judgement through debate with others.

Similar points apply in an attenuated fashion to members of communities my community regularly interacts with. Coordinating with them effectively contributes to self-maintenance in my community, and vice versa.

5. Concluding remarks: Ethical inquiry and natural history, revisited

In this chapter, I have given an account of ethical inquiry based on the organisational account of virtue. Thus, I have presented *bio-constructivism*, an account of ethical inquiry, ethical judgement and ethical truth based on the organisational account of virtue. In it, the inquirer relies on his own and his community's prior ethical judgements, but revises them according to constraints based in biology. Thus, while ethical truth is relative to an agent's prior judgements and their community, this relativism is constrained by the shared function of ethical inquiry.

Let me briefly return to the connection between ethical inquiry and 'natural history' for some concluding remarks (cf. Chapter Three). Ethical inquiry as described in this chapter resembles the form of natural history described by Michael Dove (2024), which he finds exemplified across history in Pliny the Elder, Georg Eberhard Rumphius, Carl Linnaeus, Alfred Russel Wallace, Harold C. Conklin. Such natural histories gather and present, in one form of inquiry, descriptions of the behaviours of living beings, their place in eco-cycles, human folklore about them, whether true by scientific standards or not, as well as their uses by humans. This includes agricultural, medical and ritual uses. Thus, from a modern scientific perspective, such natural histories combine objective facts about parts of the natural world as such, with facts which concern the relation of some humans to them and are thus relative to communities. With a lot of nature, there are some reflexive effects. For example, the belief that a certain plant is valuable as a spice may make it a valuable spice. This can cause it to be grown more widely, and even affect its features through artificial selection. However, their interactions are few and far enough between so that we can generally disentangle the content of such natural histories into botany, zoology, geology etc. on the one hand and sociology, anthropology, ethnology etc. on the other hand. However, with the natural history of human beings, such looping effects are so pervasive that this is not possible.

SIX: The virtues of (distant) others

In this chapter, I will describe how bio-constructivism accounts for ethical evaluations of agents who are members of distant communities. This is a prima facie challenging topic. After all, maximally functional virtue-inquiry only draws on an agent's prior judgements and those of their community's tradition because these judgements already affect their actions. If this was not the case, drawing on these judgements would not enable consistent agency or cooperation. However, this is not the case at all with members of distant communities. Hence, in inquiry concerning how to evaluate such agents, inquirers cannot draw on their prior judgements (except in counterfactual ways). Drawing on the prior judgements of the evaluee and their community, on the other hand, may seem troubling. After all, many communities have accepted horrendous practices. Are those really vicious *according to their practitioners' own traditions*?

In order to meet this challenge, I outline four approaches to such evaluations of distant others which are compatible with bio-constructivism. I reject one of these approaches, namely simple deference. The other three approaches are complementary to each other and provide an account of ethical evaluations of distant communities for bio-constructivism. In Section 1, I describe approaches to evaluating *particular* distant communities. Besides simple deference I describe two approaches, namely counterfactual inquiry and perspective-taking inquiry. Counterfactual evaluations are not strictly speaking about whether the evaluee is virtuous. Rather they are about whether someone with his or her character *would be* virtuous in a community that interacted regularly with the inquirer's community. Perspective-taking inquiry concerns their actual virtues. It starts out from the prior judgements and traditions of the evaluee rather than those of the inquirer. However, the inquiry can still rely on her capacity to follow the *logic* of ethical inquiry. Furthermore starting out from the prior judgements of the evaluee does not mean attempting to anticipate which conclusion the evaluee would most likely reach. Rather, it means inquiring into which conclusion they would reach if their capacity for ethical inquiry did maximally actualise its bare function. These two might come apart due to *cultural pathologies*: dysfunctional but self-perpetuating patterns of practices and ethical judgements comparable to the obesity/low-exercise compelled discussed in Chapter Four.

In the remainder of the chapter after that, I present ways to argue for *universal virtues* according to bio-constructivism, i.e. virtues on which all maximally functional virtue-inquiry would converge, no

matter the material conditions of an inquirer's community or their prior judgements, i.e. their *tradition*. Such external arguments for universal virtues are based on the function of virtue-inquiry as a general regulatory subsystem. Virtues are universal if their lack undermines the function of virtue-inquiry. I will first introduce the notion of such external arguments and then provide arguments for two particular sets of universal virtues, namely virtues of moderation and virtues of justice. Finally, I discuss two brief case studies of historically widespread practices which nonetheless violate both virtues of moderation and virtues of justice, viz. slavery and cruelty in war.

In discussing the virtues and the case studies, I draw on analogies to other animals, who face similar challenges which are resolved through regulatory subsystems (though not general ones). Thus, this chapter serves the secondary goal of demonstrating the epistemic fruitfulness of the organisational account of virtue. Having a category founded in biology which applies both to human virtues and the behavioural traits of other animals can produce valuable insights into either.

1. Virtue-inquiry concerning *particular* distant communities, and cultural pathologies

According to bio-constructivism, which character traits are virtues in an agent, and hence which considerations will count as reasons for them to perform which actions, depends on the conclusion they would reach in virtue-inquiry if their virtue-inquiry maximally actualised its function. Which conclusions these are depends on their own prior ethical judgements and the tradition of their community more broadly.

Given this, how can I make judgements about the virtues of agents who are not members of our community, and especially concerning people whose communities are in the past or completely causally isolated from my community? There is no way in which my prior judgements could have affected the prior actions of those agents. Hence, drawing on those judgements would not help us at all to determine which character traits would lead to future actions that are consistent with the evaluee's past actions. Similarly, the agent is causally isolated from the inquirer's community. Hence, that community plays no role in regulating the evaluee's character traits. Thus, they do not determine which character traits are functional. Hence, if my aim is to determine what traits are virtues in such

agents, I cannot straightforwardly draw on my community's tradition. After all, covering with *my* prior judgements and those of *my* community wouldn't prevent members of distant communities from undermining *their* own agency or allow them to better cooperate with *their* community.

In the following, I first discuss a strategy of simple deference in evaluating members of distant communities. I argue that this strategy is inappropriate because of the phenomenon of *cultural pathologies*. Given this, I describe two further ways to evaluate distant communities: counterfactual evaluations and evaluations based on *perspective-taking virtue-inquiry*.

1.1 Simple Deference

At first, one might think that the best way to make such judgements is to defer to dominant or consensus judgements within that community. If there is a general consensus within that community that, say, a vegetarian lifestyle is virtuous, then being vegetarian is virtuous for its members.

However, this is problematic for two reasons. Firstly, this is difficult to accept on grounds of content. There have been communities, such as those portrayed in the Iliad and in Norse sagas, which held a general consensus that cruelty in war as well as the rape and enslavement of 'outsiders' were not vicious. Instead they were taken to be indicative of an outstanding, 'excellent' or 'virtuous' life. It is difficult to accept that these communities did not get something wrong, and that the rapists and enslavers among them were not vicious.

Secondly, it is inconsistent. In inquiring for myself into which considerations to treat as reasons, i.e. in ordinary virtue-inquiry, I do not simply defer to the dominant views in my community. Rather, I proceed on the implicit assumption that the logic of virtue-inquiry may require me to adopt novel views that are not shared in my community. Thus, I implicitly assume that my community's consensus can be false. It could, of course, still be the case that deferring to the general consensus is the best we can do in evaluating members of distant communities. If so, consistency only demands that I be less certain in my deference-based evaluations of distant community members than I am of the evaluations of myself and my fellow community members that are based on virtue-inquiry.

However, firstly, such deference is still difficult to accept on first-order grounds. Should we really accept that it was *probably* virtuous for the vikings to enslave foreigners? Furthermore, the

phenomenon of '*cultural pathologies'* may give outsiders, including distant outsiders, an epistemic *advantage* in evaluating members of other communities. This advantage sometimes makes up for other disadvantages, thus making deference inappropriate. It also means that sometimes, I ought *not* to be less confident in my evaluations of distant cultures.

1.2 Cultural Pathologies

Cultural pathologies are patterns of practices and ethical judgements which are stable and self-perpetuating but nonetheless dysfunctional. Hence, the ethical judgements involved in cultural pathologies are false.

Thus, cultural pathologies resemble the anxiety- and obesity/low-exercise patterns described by Boorse (1976) and Garson (2017) in their objections to organisational and related accounts of function (see Chapters Three and Four). In the obesity/low-exercise example, an agent's obesity makes exercising more difficult and frustrating. Thus, they exercise less. This maintains their obesity. Similarly, a clinically anxious agent notices more potentially threatening features in their environment, hence maintaining their high level of anxiety. Pre-theoretically, the anxiety and the obesity are *dys*functional traits. However, according to the critics, organisational and cybernetic accounts of function imply that they are functional. This is because they found function on self-maintenance.

However, the organisational account can explain why they are not functional. Firstly, they do not contribute to the complex web of constraints that constitutes the organisation of the organism. With a functional trait, every other trait within the organisation is maintained by it in some direct or indirect way. For example, my spleen filters aged red blood cells from my blood. In doing so, it improves my blood's ability to carry oxygen, nutrients, hormones, etc. Most of my other functional traits depend on this. In contrast, the obesity maintains only the low level of exercise, and the low level of exercise maintains only the obesity. The anxiety maintains only itself and the high perception of threat. Secondly, the obesity and the anxiety are not regulated-for by any regulatory subsystem of the organism. Plausible, they are regulated *against* by some subsystem.

Cultural pathologies are similar isolated self-perpetuating patterns within communities which interfere with the functioning of the members' capacities to treat considerations as reasons. Thus, cultural pathologies also resemble "mind-viruses" as described by Dawkins (1993). Mind-viruses are 'memes', or bits of self-perpetuating information, which are disposed to spread even though they harm their 'hosts', i.e. the human agents enacting them. Unlike Dawkins, my emphasis is on the persistence of dysfunctional practices and judgements in a community rather than on their spread. I am also not committed to using the same notion of *function* as Dawkins or his memetic theory.

Many cultural pathologies by that definition relate to an apparent advantage which pushes the community members to pursue selfish motivated reasoning and exclude some members of their economic community from shared ethical inquiry, but which actually undermines virtue-inquiry.

Consider, for example, a community that practices slavery but relative to whose tradition, enslavement is vicious. In this community, let us suppose, the judgements that enslavement is permissible are inconsistent with many other prior judgements such that the best way to make the judgements consistent would be to conclude that enslavement is impermissible. Hence, the pro-slavery judgements violate the consistency constraint (cf. Smith 2021). Thus, they undermine the agents' ability to pursue goals consistently, as well as making their behaviour less predictable to others. They also violate shareability constraints. Thus, the enslavers cannot collaborate with the enslaved in ethical inquiry. They cannot benefit from their experience and practical knowledge in their ethical deliberations. They also cannot expect the enslaved in turn to share their ethical inquiry with the enslavers. Hence, they must fear theft, sabotage and poisoning from their enslaved labourers and servants.

In such a community, plausibly, most enslavers and other free members would judge that enslavement is permissible. Given their daily experience, slavery will appear to be strongly in their interest. After all, it allows the enslavers to avoid physical labour; even free community members who do not own slaves can avoid the most arduous and degrading work. Furthermore, enslavement may entail psychological benefits (cf. Alcoff 2006). Free members of a slave-owning society will see themselves as superior to the enslaved. As human inquirers, we are biased to engage in motivated reasoning. We are disproportionately likely to reach conclusions which we believe it is in our interest to reach, even if these conclusions are not supported by our evidence and the logic of the inquiry we

are engaging in (cf. Mercier and Sperber 2017). Notoriously, Thomas Jefferson (2022 [1783]) takes seriously the thought that Black people's lack of flowing hair indicates that they are naturally inferior to whites. Thus, free members of communities practicing slavery are likely to reach the conclusion that slavery is permissible, whether or not this is the case relative to their tradition.

Such motivated reasoning is relatively easily prevented if others, who do not share the same apparent interests, evaluate one's reasoning. However, in the slave-owning community, virtually all free members have some apparent interest in concluding that slavery is permissible. Thus, they are less likely to challenge the motivated pro-slavery conclusions. The enslaved population would, of course, be much more likely to criticise pro-slavery arguments. However, as part of the practice of slavery, the enslavers and most other free members dismiss the ethical arguments of the enslaved. It is, in fact, perfectly reasonable for the enslavers to assume that the enslaved would use ethical debate to manipulate them rather than engage in 'genuine', good-faith ethical debate.

Furthermore, through the conditions of enslavement, the enslaved will be prevented from devoting much time and energy to ethical inquiry and from communicating with each other to engage in collaborative ethical inquiry. They can also not easily conduct 'experiments in living' which would give them more empirical evidence for ethical inquiry. Thus, while the enslaved will know that it is not vicious for them to try and escape slavery, it is more difficult for them to come to know that enslavement as such is vicious. Uprisings by enslaved people in antiquity seemed to have aimed at attaining freedom for their participants, but not for all enslaved (cf. Williams 1993). Formerly enslaved people from, say, the ante-bellum South or Haiti/Saint-Domingue produced powerful universal attacks on slavery (consider e.g. Frederick Douglass 2015 [1845] and the authors of the 1804 Haitian Declaration of Independence [see Armitage and Gaffield 2016]). However, most if not all of these thinkers and activists had biographical connections to communities where slavery wasn't practiced at the time: Europe, the Northern states, and many of the indigenous communities enslaved Africans had grown up in, etc. Thus, in isolation, the complex consisting of pro-slavery judgements and the practice of enslavement will self-perpetuate. Even if exposed to non-slave-owning societies, it will be relatively stable.

Looking out for such potential cultural pathologies is a good *heuristic* for constructing arguments in ethical inquiry and avoiding biases. It is such a heuristic both in ordinary ethical inquiry, directed at

what to do oneself, and at inquiry aimed at evaluating others. However, it does not itself provide such arguments. Rather, it must be established through an autonomous ethical argument that a certain practice is vicious.

Due to cultural pathologies, outsiders, including distant outsiders, are at least in some respect in a better position to inquire into the virtues of a community's members than those members themselves. Insiders, but not outsiders, are affected by the community's cultural pathologies. Of course, this advantage is offset by disadvantages. It will be more difficult for outsiders to understand the tradition of the community, and to have empirical knowledge of their material conditions. Nonetheless, the existence of cultural pathologies counts against an epistemic policy of simple deference concerning virtues in distant communities.

1.3 Counterfactual Virtue-Inquiry

One might also adopt a fictionalist position concerning evaluations concerning virtues in distant communities. According to such an account, if I am evaluating distant community members, I am evaluating them either *as if* they were members of my community, or *as if* our communities were to become conjoined. In the first case, I would rely merely on my own community's tradition. In the latter, I rely on both the tradition of my own community and on theirs. It may indeed sometimes be the case that we engage in such counterfactual inquiries if we are thinking about whether members of such distant communities are virtues. After all, such counterfactual inquiries may be much more useful to us than any inquiry into the actual virtues of the distant communities. We will never actually interact with that community. Hence, what would be the benefit of knowing what traits are virtuous for them? In contrast, the counterfactual inquiries may serve as useful practice for evaluating strange agents in our own community and as sources of analogies for ordinary ethical inquiry.

I believe that it is plausible that some of the occasions on which we evaluate distant community members are best understood on such a conception. Suppose we are debating whether Tiberius Gracchus was a progressive, enlightened and compassionate champion of egalitarian justice or a populist authoritarian rabble rouser. Plausibly, we are applying contemporary virtue-concepts to him, which have little basis in his own historical context. Of course, this does not mean that such debate

is useless. This may plausibly explain some of our most seemingly confident ethical evaluations of distant communities.

However, it is unacceptable on first-order grounds that *all* evaluations concerning distant communities are best construed like this. Ante-bellum slave-owners were *actually* unjust, not just such that they would have been unjust if they lived among us, today. Furthermore, it does not acknowledge that the existence of cultural pathologies means that outsiders sometimes have an epistemic advantage in evaluating distant communities. This is an advantage in inquiring into what is *actually* virtuous for members of that community, not concerning what *would be* virtuous if they were members of one's own community.

1.4 Perspective-taking virtue-inquiry

Besides simple deference and counterfactual inquiry, I suggest that an especially important method for evaluating virtues in other communities consists in *perspective-taking ethical inquiry*, including *perspective-taking virtue-inquiry*. This inquiry is *perspective-taking* in the sense that the inquirer aims to 'put themselves in the shoes' of the agent they are evaluating. Rather than relying on their own prior judgements and the tradition of their community, they rely on the prior judgments of the agent who is the target of evaluation and the tradition of his or her community. They also do their best to consider the material conditions of the community in anticipating the feasibility of living according to a given conception of virtue and its effects on the agent's thriving (cf. Chapter Six). By the material conditions, I mean, roughly, their natural environment and their available technology. Thus, 'chastity' may be a more important --- and more restrictive --- virtue in a community where contraception and treatment for sexually transmitted diseases are severely limited and where most of the essential economic activity consists in agriculture on small farmsteads that can most be operated most effectively by family units, such that certainty of inheritance is important to ensure continuous economic activity.

Thus, in perspective-taking virtue-inquiry, the inquirer starts out with the judgements of the community whose members' virtues she is inquiring into. She revises these judgements in line with the logic of ethical inquiry. Thus, she revises inconsistent judgements as well as judgements that are incompatible with thriving or not shareable, etc. By reasoning in line with the feasibility and thriving constraints, she relies on her empirical knowledge of the target community's material conditions. As

with other forms of inquiry, such perspective-taking inquiry will often benefit from being pursued collaboratively with co-inquirers. In conducting this perspective-taking ethical inquiry, the inquirer(s) will draw on expertise and skill they have developed through ordinary ethical inquiry.

Perspective-taking ethical inquiry is not the same as anticipating which judgements members of the target community would typically or most likely make. Those could, for example, be judgements implicated in cultural pathologies. Rather the inquirer aims to determine which judgements members of the target culture would make if their capacity for virtue-inquiry was fully actualising its bare function.

Perspective-taking virtue-inquiry is possible because virtue-inquiry has the same bare function qua general regulatory subsystem in any human being as long as it acts as a general regulatory subsystem at all. This is because as a general regulatory subsystem it must make a systematic and consistent contribution to self-maintenance *as such*, i.e. one which applies to all levels of self-maintenance. Such contributions are rare. Virtue-inquiry makes such a contribution by allowing us to pursue goals effectively without self-sabotage and while cooperating with fellow community members. There is no possible contribution to self-maintenance that is *similar* to this, but not the same. There may be other logically possible general contributions to self-maintenance. However, if so, these would have to be substantially different from the contribution actually made by paradigmatic virtue-inquiry. Given the empirical similarities between all human beings, no human beings are likely to have a general regulatory subsystem with such a hypothetical function.

In other words, virtue-inquiry fully actualising its bare function is a 'peak' or local maximum in the landscape of possible general contributions to self-maintenance. It is not close to any other such peaks. Hence, any faculty that is even just somewhat similar to maximally maximally functional virtue-inquiry would increase its general contribution to self-maintenance by becoming closer to such virtue-inquiry. Hence, all instances of virtue-inquiry in human beings have the same bare function, which they imperfectly actualise. Since the logic of virtue-inquiry is determined by that bare function, it is the logic of all instances of virtue-inquiry. Hence, any human agent can rely on their capacity for (ordinary) virtue-inquiry to engage in perspective-taking virtue-inquiry concerning the virtues of agents in distant communities.

2. External arguments for universal virtues

In the last section, I have argued that due to the shared bare function of virtue-inquiry, human beings can engage in perspective-taking virtue-inquiry. In the remainder of this chapter, I argue that there are universal virtues on which all maximally functional virtue-inquiry would converge, no matter the inquirer's particular prior judgements or their community's tradition and material conditions. Thus, such virtues are virtues for all human agents. While the broad outline of these virtues are shared and there are certain behaviours that are contrary to them for all human agents, the 'substance' of the virtues are not shared between communities. Thus, even though in every community, there will be virtues of justice, and they will exclude certain practices, such as enslavement, what it takes to be just will differ from community to community. These implications coincide with recent results concerning 'folk metaethics'. Lay participants in studies seem to treat some central ethical judgements as universal but many others relativistically (e.g. Poelzler and Wright 2020).

Such universal virtues can be demonstrated through *external arguments*. These are arguments which show that, in order to maximally actualise its bare function, virtue-inquiry must support certain conclusions. Any instance of virtue-inquiry that did not support these conclusions would not be maximally functional. Such external arguments cannot replace arguments internal to virtue-inquiry internal to the particular tradition. Firstly, such internal arguments are necessary to determine the substance of the virtue in question more precisely. Secondly, such *meta-inquiry* into virtue-inquiry is incapable of regulating character traits in the way actual virtue-inquiry is. This is because virtue-inquiry presents its conclusions under a special mode of presentation, which is linked to motivation for action (see Chapter Seven). Hence, the external arguments only show that internal arguments following the autonomous logic of virtue-inquiry *must be* available if virtue-inquiry is to actualise its distinctive function as a general regulatory subsystem. The meta-inquiry containing them cannot take the role of virtue-inquiry as a general regulatory subsystem.

It is at least logically possible that some traditions are so corrupt and ridden with cultural pathologies that relative to them, the conclusions demanded by the external arguments could not be reached. If so, apparent virtue-inquiry in such communities would not make a contribution to the self-maintenance as such. Any apparent contribution to self-maintenance as such would be

continuously undermined by itself. In such communities, ethical evaluations of individual agents would be meaningless. Alisdair MacIntyre's (1981) pessimistic argument in *After Virtue* can be read as suggesting that the communities of the modern, capitalist West (and most of its [former] colonies) are in such a predicament. However, given the extent to which we can recognise the universal virtues argued for in this chapter, I believe that bio-constructivism would suggest that such a conclusion might be hyperbolic.

2.1 Analogies with other animals

In arguing for universal virtues, I draw on analogies between behavioural regulation in humans and in other animals. Humans and other animals face many of the same challenges, which may be resolved in ways that benefit different levels of self-maintenance. With human virtue, the favoured resolution is determined by ethical inquiry. The precise way in which it is resolved depends on the particular tradition of a community and its material conditions. However, the favoured resolution could not be one that undermines the function of virtue-inquiry and virtue-inquiry itself. Hence, the challenge corresponds to universal virtues, which must be recognised by all maximally functional virtue-inquiry.

Other animals lack a general regulatory system like virtue-inquiry. Thus, insofar as there is any typical or 'favoured' resolution, it is the resolution which maximises inclusive fitness. For human beings, the resolution actualising the relevant regulated function, and thus --- according to the theory I am advancing --- the virtuous resolution, is the resolution that would be endorsed by maximally functional virtue-inquiry. This may or may not be the resolution that maximises inclusive fitness. After all, for the organisational account of virtue, different levels of self-maintenance are incommensurable (see Chapters Four and Five). It may be that selection pressures have shaped virtue-inquiry and our prior judgements such that, if ethical inquiry functions well, it will always end up endorsing the fitness-maximising resolution. However, if so, this can only be discovered by engaging in ordinary and perspective-taking virtue-inquiry and comparing the conclusions reached to the resolutions predicted by considerations of inclusive fitness. Since the relevant notion of function does not privilege the lineage-level, this could not be derived through meta-inquiry about virtue-inquiry.

My methodology in this chapter thus resembles Nussbaum's (1988) approach in "Non-Relative Virtues". Nussbaum discerns spheres of experience in human life and identifies virtues with the proper attitudes towards such spheres. For example, a universal experience in human life contains potential dangers to one's life and body and other things one values. The proper attitude towards such dangers is *courage*. Thus, the reference of 'courage' is fixed, whatever courage might exactly consist in. However, unlike Nussbaum, I attempt to identify problems in the lives of wider classes of animals and/or living beings in general. The human virtues constitute one way to resolve these conflicts.

Another precedent for my approach is MacIntyre's (1999) *Dependent Rational Animals*. MacIntyre argues that there are certain virtues which human beings (have and) need due to belonging to a larger class, i.e. that of intelligent social mammals. These are the *virtues of acknowledged dependence*. These are virtues like charity, which are needed for a species of social and vulnerable, dependent animals to survive. Mary Midgley's (1979) *Beast and Man* pursues a similar project in connecting "human nature" to the "natures" of other animals. For example, she compares the role of human morality with the purported role that "aggression inhibition" plays in the lives of other animals. Unfortunately, both MacIntyre and Midgley rely on outdated descriptions of animal behaviour, which are influenced by naive group-selectionism. Thus, they are happy to identify behaviour that is 'necessary for the survival of the species', such as allo-parenting and collaborative care for the young, without considering *how* such behaviour is stably maintained despite the conflicting evolutionary interests of the participants. Of course, neither MacIntyre nor Midgley identify virtue with evolutionary function of any kind, group-selectionist or otherwise. However, the systematic description of animal behaviour which they rely on have been influenced by naively group-selectionist assumptions.

In the following, I will consider two broad categories of such virtues. The first are *virtues of moderation*, the second are *virtues of justice*

3 Virtues of moderation

One class of virtues that must be recognised by all maximally functional virtue-inquiry can be called *virtues of moderation*. These are virtues which human beings need to keep self-reinforcing feedback loops involved in motivation in check. It is a common observation that virtues are opposed to "vices" or "temptations". For example, Foot (1978: 9) claims that "[t]he virtues are corrective, each one standing at a point at which there is some temptation to be resisted or deficiency of motivation to be made good." With the virtues of moderation, this counteracting or regulating aspect is especially evident. They are distinguished from other virtues by being opposed to biologically basic motivational systems, which we might call *drives*. These include hunger/appetite, sexual attraction, fear, anger and the inclination towards rest. These motivational states promote (inclusive) fitness in direct ways. Thus, they are likely to have deep evolutionary roots. It is also likely that many other animals will have analogous motivational subsystems, whether or not these are phylogenetically related. Given the direct contribution such subsystems make to self-maintenance, it is likely that they were selected to exert a strong motivational force. In humans, they can be experienced as strong, even overwhelming emotions or feelings.

Plausibly, the drives involve positive (or *self-reinforcing*) feedback loops (see e.g. Nagel 1969, Adam and Epel 2007, De Jong 2009, Kimble at al. 2014, Hollenstein 2015). An episode of fear, for example, begins with me noticing a small feature of my environment indicating danger. In response to this feature, I begin to notice more features of my environment that may indicate danger. I also begin to think about potential ways I might be in danger that are not indicated by my environment. Some bodily responses to fear are triggered, such as increased heart rate and breath frequency. I notice these bodily changes. All this makes me more fearful. These feedback loops also affect my behaviour in the long-term. They are habit-forming: they entrench connections between different components of the cognitive systems. Thus, if I regularly experience such episodes of fear, I will grow more fearful.

Such self-reinforcing feedback loops are effective in producing motivation in situations where doing so is of great importance for the survival of the individual or the lineage. However, they also carry the risk of 'running away' and overwhelming all other motivational systems. Thus, there is a need for regulatory systems that keep such feedback loops in check.

If the self-reinforcing feedback loops are not kept in check, they will hinder the self-maintenance of the agent on all levels. They will hinder it directly by leading to various behaviours that harm self-maintenance on all levels, such as over-eating in the case of appetite, and hyper-vigilance or panicked flight in response to a relatively minor threat in the case of fear. Crucially, it also undermines our ability to regulate our behaviour through ethical inquiry. Thus, it interferes with a capacity that benefits self-maintenance on all levels. Hence, keeping self-reinforcing motivational feedback loops in check is necessary for ethical inquiry to maximally actualise its bare function. Hence, the virtues involved in it must be recognised as virtues by maximally functional virtue-inquiry in all communities.

In human beings ethical inquiry is one important system that may disrupt such positive feedback loops. I may judge 'I am being a coward'. This judgement disrupts the feedback loop, either through a direct exercise of willpower or by motivating me to use more indirect strategies of self-control. I may, for example, articulate a self-command ('Stop being such a yellowbelly!'), or redirect my attention either to the expected positive outcomes of acting courageously or to factors that make acting courageously relatively safe, despite appearances.

Ethical judgements are not involved in all the ways in which the positive feedback loops are disrupted. Instead, in the fully virtuous agent, the feedback loops will come to a stable halt without the intervention of ethical inquiry. Instead, there are mechanisms that may disrupt such feedback loops. These mechanisms may be in part the result of habit-forming through ethical self-control, otherwise learned or innate. In the ideally virtuous agents, the feedback loops are disrupted by these mechanisms. Ethical inquiry does not need to intervene. Thus, we can say that the virtues of moderation are constituted by these mechanisms.

Many other animals face the same challenge of regulating such motivational feedback loops. After all, there are different, possibly conflicting goals, in the lives of such animals. They must avoid danger and exertion, find food and mates, care for their offspring, etc. Many of these goals have dedicated motivational systems associated with them. In order to balance the goals, these dedicated motivational systems must be regulated. Thus, many animals have motivational systems that are analogues to the systems constituting the virtues of moderation in humans.

The virtues of moderation can be described using a 'vice-first' method. Based on empirical knowledge about human beings, including knowledge derived from biological science such as evolutionary psychology, we can identify self-reinforcing feedback loops that are prone to get out of control without checks on them. The virtues of moderation are the vices opposed to such vices.

The 'vice-first' approach is related to but different from Aristotle's 'doctrine of the mean'. Aristotle's approach, as reconstructed by Nussbaum (1979) begins by isolating spheres of human experience, such as "fear of important damages, esp. death" or the "distribution of limited resources". Virtues consist in the appropriate attitudes towards the spheres, i.e. the attitudes that are best for humans. For Aristotle (though not necessarily for Nussbaum), this attitude will always be the mean between two extremes. For example, courage, the proper attitude as regards fear of important danger, is the 'mean' between rashness and cowardice. Calling it the 'mean' does not imply that it is an average or median in any algorithmically determinable sense. It only means that it can be contrasted with both a vice of deficiency and a vice of excess.

Unlike the method derived from the doctrine of the mean, the 'vice-first' method gives priority to *one* vice contrasting with the virtue, i.e. the one related to a run-away positive feedback loop. Nonetheless, the 'vice-first' method is compatible with the method of the spheres and even with the doctrine of the mean. The virtues of moderation concern a particular kind of sphere of experience. These spheres are each defined by a particular, basic motivational system which is prone to harmful positive feedback loops. Regularly allowing it to enter such feedback constitutes one vice the virtue contrasts with. There will be a second vice this virtue could be contrasted with. However, this second vice is secondary. Often, it consists of the regulation of the first vice becoming excessive. For example, the virtue of courage regulates fear. Primarily, it is opposed to cowardice, an excessive feedback loop of fear. However, it is also opposed to recklessness. One form of recklessness consists in overcompensating for fear. An agent who is reckless in this way experiences fear. They 'over-regulate' their fear, and thus take unnecessary risks. For example, they may take part in dangerous 'dares' simply to prove that they are not cowards.

There may also be such vices of 'deficiency' that do not consist in such overregulation. For example, some agents may really be reckless because they experience insufficient fear. In some cases, there

may be other virtues that are primarily opposed to these secondary vices. For example, whereas courage is primarily opposed to cowardice and secondarily to recklessness, caution is primarily opposed to recklessness. Caution and courage are complementary virtues. Such virtues may themselves be virtues of moderation, if they are opposed to a different basic motivational system. In other cases, this 'deficiency' may be beyond the influence of ethical inquiry. If so, it is not a defect of character. It may not be a defect at all. For example, an agent whose sexual desire is overregulated manifests a vice. An agent who is asexual and has little sexual desire without over-regulation does not.

3.1 The 'Lively Virtues'

The exact application of the 'vice-first' methodology of arguing for universal virtues depends on empirical knowledge. However, one set of plausible virtues of moderation is provided by the list of 'lively' or 'remedial' virtues of Catholic tradition. This list was first composed by Pope Gregory I and expanded by Thomas Aquinas (1991 [1274]). The seven vital virtues are each opposed to one of the seven deadly sins. These 'deadly sins' can be seen as self-reinforcing motivational feedback loops running out of control. The list was at least in part based on empirical experience of human beings. Its elements have precursors among the virtues recognised by Aristotle and analogues among virtues recognised in other traditions, such as the Dharmic traditions.

The seven deadly sins and lively virtues according to Thomas Aquinas are: lust and chastity, gluttony and temperance, greed and charity, sloth and diligence, envy and kindness, wrath and patience as well as humility and pride. Most, perhaps all of these plausibly regulate some evolutionarily basic motivational system. Chastity regulates sexual desire, temperance appetite, especially in so far as it is associated with 'opportunistic' rather than need-driven eating, and the desire for sensual pleasures more generally. Charity regulates the desire for the acquisition of goods; diligence the desire to be and remain at rest. Patience in the Thomistic senses regulates the motivational system involved in aggression against conspecifics. Humility regulates status-seeking behaviour.

With most of these motivational systems, it is very likely that human beings share them with other animals. Thus, it is also plausible that they involve distinctive self-reinforcing feedback loops. With envy and kindness, this is somewhat less clear. Introspectively, it is plausible that envy, too, can

involve self-reinforcing motivational feedback loops. One can get all worked up in envy about someone else's good fortune. This suggests that there is a use for virtues regulating envy. De Waal and colleagues (e.g. de Waal and Sherblom 2018) argue that non-human primates experience envy, and that envy is the evolutionary basis of human systems of justice. However, his findings have been contested (e.g. Tomasello and Vaish 2013).

Most of the lively virtues are still recognised as virtues in Western societies. However, for some items, this is more controversial. Firstly, it may appear strange for there to be a vice opposed primarily to gluttony. Fat activists have rightly criticised the association between fat bodies and weak self-control. However, firstly, body type is determined by many factors besides eating habits. Thus, one can recognise a vice of 'gluttony' without treating body type as a major indicator thereof. Secondly, the mediaeval vice of gluttony had more to do with the harm the glutton causes to others through waste than the harm they cause to themselves. We are still concerned with the effect our food consumption and other sensory pleasures have on our fellow human beings, e.g. through climate change, waste of water resources, etc.

Similarly, one might think that we do not consider 'chastity' a virtue anymore. However, we still recognise the need to regulate sexual desire. Imagine someone who regularly thinks of many of their co-workers as potential sexual partners. This interferes with their ability to perceive their skills as co-workers, etc. We would think of such a person as having a vice. In modern vocabulary, we may call them a *creep* and their attitude *objectifying*. In a more old-fashioned register, you may still call them *lecherous*. There must be a virtue opposed to such vices. There is no fully established name for it, but we may still call it *chastity*. Such usage has precedent, for example, in Nietzsche. He observes: "There is no necessary contradistinction between sensuality and chastity, every good marriage, every genuine love affair is beyond this contradistinction" (Nietzsche 2008 [1889]: 71, translation adjusted) and "Chastity a virtue with some, but with many almost a vice" (Nietzsche 2012 [1883]: XIII). By 'chastity' (*Keuschheit*) he seems to mean a kind of self-restraint in sexual matters. For him, it is virtuous if it is guided by authentically affirmed values.

Finally, we may not recognise *humility* as a virtue. However, we still recognise that status-seeking behaviour can become viciously excessive. We condemn people as *arrogant*, *vain* or *prideful* for

engaging in vicious status-seeking and status-asserting behaviour. There must be a virtue opposed to such vices, even if we are more likely to characterise it as, say, *appropriate self-regard* rather than *humility*.

Analogues of the lively virtues have been recognised in historically and geographically distant communities. Aristotle recognises *temperance*, which is opposed to *self-indulgence* as a vice of excess. Unlike the later Catholic lists, Aristotle does not single out self-restraint in the particular domains of food and sexuality. *Greatness of soul* and *proper ambition* correspond to *humility/wholesome self-regard* in that they are opposed to *vanity*. *Gentleness* is opposed to *wrath/irascibility* (cf. Rodriguez 2014), and *righteous indignation* (on one's own behalf) to *envy*. *Liberality* as the appropriate attitude towards one's possessions and limit on greed and miserliness. Thus, it corresponds to charity. Aristotle does not recognise a particular virtue concerned with sexuality; however, it can be seen as included in temperance.

The *Shandilya Upanishad*, written around 100-300 BCE in India, contains a list of 10 *yamas*: 'virtues' or 'virtuous disciplines'. Among them are *ahimsa* 'non-violence', analogous to gentleness/patience; *brahmacharya* 'being like a celibate', which refers not to literal celibacy but to temperance or self-restraint with a special focus on sexual matters and is thus analogous to chastity and temperance; *aparigraha* 'non-possessiveness', analogous to charity and to kindness in its opposition to envy; *dhṛti* 'perseverance', analogous to diligence and *mitahara* 'measured diet', a virtue that can be seen as specifically opposed to gluttony.

The list of the Shandilya Upanishad contains no direct analogue to humility or greatness of soul. However, the later Bhagavad Gita (13.07) refers to humility (*amānitva*), pridelessness (*adambhitva*) and absence of false ego (*anahankarah*) as virtues alongside steadfastness (*sthairyam,* cf. diligence), self-control (*atma-vinigrahah*), "renunciation of the objects of sense gratification" (*indriyarthesu vairagyam*, cf. temperance and chastity) and "nonattachment to … home etc." (*asaktir anabhisvangah … grha-adisu*, cf. charity/liberality).

3.2 Courage

Courage is not included in the traditional Catholic list of the seven lively virtues (though fortitude, which is closely related, is included in the separate list of *cardinal* virtues). Nonetheless, courage can similarly be understood as a virtue regulating *fear*. Fear is a motivational system with deep

evolutionary roots. Most mammals show fear in a way that is recognisable to us. Fear indicates dangers to the survival of the individual, or (at least in the case of humans) other goods they care about. Fear may easily lead to positive feedback loops. Such positive feedback loops can lead to behaviour that harms self-maintenance on all levels. Thus, both humans and other animals have a need for regulatory systems that can disrupt such feedback loops. With human beings, excessive fear would undermine the ability of fear itself and these regulatory systems are ultimately regulated by ethical inquiry. Thus, for human beings, there is a virtue of *courage*. Since courage is ultimately regulated by ethical inquiry, the amount to which fear ought to determine one's actions depends on how important one would judge different goals to be given that one's faculty of ethical inquiry is well-functioning. Thus, the courageous person is prepared to accept much greater danger to save a life than, say, to win a car in a competition.

As mentioned above, there is plausibly another virtue, *caution*, which is opposed to being insufficiently fearful. Thus, caution is one of the virtues that regulates the several other motivational systems that may lead us to inappropriately disregard danger. Being cautious is a component of being temperate, chaste, diligent, patient, etc.

4 Virtues of justice

Another class virtues are the virtues of justice. These are virtues resolving dilemmas that are structurally analogous to a *tragedy of the commons*- Such dilemmas occur for many animals and other living beings (Rankin et al. 2007; cf. Hardin 1968). In the thought experiment providing the conventional name for such dilemmas, there is a common grazing land which is accessible to several pastoralists. If each pastoralist lets their full herd graze on the commons, the grass will die. Thus, for the use to be sustainable, the pastoralists would have to restrict their use of the commons. However, the actions of a single pastoralist will virtually never make the difference between the grass surviving or dying. Thus, letting their whole herd graze appears always to be the better choice for any pastoralist. If the other pastoralists are practicing restraint, then she may benefit more by letting her whole herd graze. However, if the other pastoralists are also letting their whole herd graze, she might just as well go for what benefit she can get from the dying grass and let her whole herd graze. A dilemma is structurally analogous to the tragedy of the commons if there is benefit (e.g. in the form of contributions to self-maintenance or inclusive fitness) which can be gained if a sufficient number

of individuals cooperate. However, each individual on their own can always gain a greater benefit if they defect. Situations potentially leading to such tragedies occur for humans and other animals. Often, there are regulatory systems involved in preventing them from escalating into the breakdown of the practice. With humans, these regulatory systems constitute virtues, i.e. the virtues of justice

We can distinguish two types of tragedies of the commons. In the 'external resource' type, there is an external resource which provides the benefit, and which is liable to be overexploited. The original tragedy-of-the-commons thought experiment is of this type. With non-human organisms, such tragedies of the commons occur in host-parasite interactions (Frank 1996, de Roode et al. 2005, Kerr et al. 2006) and in male competition for females (Rankin and Kokko 2006). With human beings, virtues related to this type of the tragedy of the commons may be called virtues of environmental justice. They are concerned with the preservation of the environment necessary for human thriving in the broadest sense. Secondly, there is the 'social goods' type. In this type, the potential benefit is merely produced through cooperation. There is no external resource to be exploited. Most virtues of justice related to this type of tragedy of the commons.

Some virtues of justice are universal virtues: all maximally functional instances of virtue-inquiry would converge on recognising some overlapping variants of this virtue. This is the case if the failure to resolve the tragedy of the commons hinder effective agency, i.e. the ability to pursue goals effectively. If so, preserving the good contributes to self-maintenance on all levels and thus contributes to the bare function of virtue-inquiry. It also contributes to the actualisation of any other virtue which maximally functional inquiry might recognise relative to any tradition. This is because these virtues depend on effective agency. In the following, I will consider four examples of such virtues, i.e. promissory fidelity, charity, ethical respect and righteousness, i.e. participation in the third-party enforcement of norms

4.1 Promissory fidelity and promise-keeping

One universal virtue of justice is promissory fidelity, i.e. the virtue involved in keeping promises. Promise-keeping is structurally analogous to the tragedy of the commons. It is beneficial for human beings to be part of a community in which promises are generally kept. The practice of promising and promise-keeping allows agents to rely on another, and thus enables various forms of division of labour. This benefits their ability to reach any goal they might want to reach. Hence it contributes to

their effective agency, and to self-maintenance at any level. In contrast, in a community without promise-keeping or an equivalent practice, the agency of community members would be severely restricted. They could not rely on other community members, and not divide their labour in efficient ways.

The practice of promise-keeping cannot exist without a virtue of promissory fidelity. Even though all agents benefit from the institution of promise-keeping, each individual agent could seemingly sometimes benefit more from breaking a promise. Their individual act of infidelity would be very unlikely to lead to the collapse of the practice. Hence, all agents will regularly be tempted to break promises. There will be strong motivational systems pushing them to break their promises. In order to keep such motivational systems from determining action, there must be some counteracting motivational systems. These are ultimately regulated by virtue-inquiry. Thus, they are virtuous.

This shows that all instances of maximally functional virtue-inquiry would lead agents to *publicly* endorse promissory fidelity. The practice of promise-keeping contributes to effective agency while its absence harms it. A virtue of promissory fidelity is necessary for the practice of promise-keeping, and that virtue can only exist in a community if it is generally publicly endorsed. Endorsing it publicly individually comes at little cost, and it contributes to the practice of promise-keeping.

If a practice of promise-keeping is not in place (if there be such communities), then maximally functional virtue-inquiry would promote it and lead agents to attempt to bring it into being. Of course, it may be easier to excuse someone for failing to do so than for failing to keep their promises if the practice is already in place.

Even if maximally functional virtue-inquiry leads agents to publicly endorse promissory fidelity, one might worry that they might still *privately* judge that they should not be faithful to their promises if keeping them is inconvenient. However, there are two reasons why this is not the case.

Firstly, there are 'enforcement mechanisms'. Given that promissory fidelity is generally publicly endorsed, these enforcement mechanisms reduce the gap between promissory fidelity and what's 'in an agent's interest'. Many of these enforcement mechanisms are reputation-based. Agents who know that another agent has broken a promise are less likely to cooperate with that agent. This is in part

because they would have to expect that they themselves will be betrayed. However, humans even forego other advantages to punish others who have broken promises (or violated norms in other ways). They are also likely to share the knowledge that an agent is a promise-breaker with others. In contrast, an agent who keeps promises even under difficult circumstances acquires a reputation that makes her an especially desirable partner for cooperation. Hence, they have more opportunities to cooperate with desirable partners themselves. Enforcement mechanisms exist not only among human beings, but also among other organisms that face tragedies of the commons. For example, among *Melipona* honey bees, eggs laid by non-queens are eaten by other worker bees. Thus, the inclusive fitness of the reproducing non-queens is reduced (Wensleers and Ratnieks 2004).

Of course, an agent may attempt to gain all of these advantages by keeping their promise-breaking secret. However, firstly such a strategy is risky. There is always a risk of being discovered. Furthermore, through positive feedback loops, an agent who begins to break promises becomes more and more likely to break promises in risky ways. Most importantly however, as described in Chapter Five doing so requires keeping one's intentions secret. They must be kept secret not just from the target of infidelity, but from virtually all others. Otherwise, the promise-breaker would acquire a bad reputation. Thus, the promise-breaker would not be able to participate in socially coupled ethical inquiry. Thus, their ability to maintain effective agency through ethical inquiry would be severely limited. Hence, given that promissory fidelity is generally publicly endorsed, maximally functional virtue-inquiry will also conclude *in private* that one ought to be faithful to one's promises. Hence, promissory fidelity is a universal virtue.

4.2 Charity and helping the needy

Another practice that produces a social good consists in more fortunate community members helping the (most) needy. In part, this is closely analogous to the social goods discussed above. After all, helping the most needy maintains social peace. If they are not supported, it may become likely (and virtuous!) for them to act against the prima facie prescription of other virtues, and e.g. employ violence and violate the existing property order. Such social peace rambles agents to pursue their goals.

Additionally, even if agents are not guaranteed to benefit from this practice by being a recipient, there is a *chance* that they will. Given marginal utility the advantage which the recipients receive is much greater than the disadvantage which the givers suffer. This advantage will benefit effective agency, as I can use the resources I receive when in need to pursue whatever goals I may have. Thus even if you are considerably more likely to be a giver than a recipient, the expected value you receive from the practice is positive. Hence, all maximally well-functioning instances of virtue-inquiry will publicly endorse charity. Due to reputation-based enforcement mechanisms and the advantages of being able to participate in shared ethical inquiry, this means that agents whose capacity for virtue-inquiry is fully actualising its bare function will also judge in private that they ought to help the needy. The virtue of justice related to the tragedy of the commons arising from the institution of helping the needy overlaps largely with the virtue of moderation keeping the acquisitive drive in check. Hence charity is both a virtue of justice and a virtue of moderation.

4.2 Ethical respect and joint ethical inquiry

Another practice that contributes to effective agency is joint ethical inquiry potentially involving all people one interacts with. Firstly, as with most forms of joint inquiry, the more perspectives participate in ethical inquiry, the more likely its participants are to discover serious errors. This is so especially since, according to bio-constructivism, ethical inquiry has an empirical component. Different participants have different empirical knowledge. Since ethical inquiry contributes to effective agency, avoiding errors ethical inquiry contributes to effective agency. Thus, it contributes to self-maintenance on all levels and to the actualisation of any virtue that might be recognised by maximally functional virtue-inquiry in any tradition. Secondly, joint ethical inquiry contributes to trust between different members of the community. Hence, they can cooperate without having to expect being betrayed. As with promise-keeping, this enables them to pursue their goals more effectively.

In order for joint ethical inquiry to be sustainable within a community, members of that community must, firstly, not generally use joint ethical inquiry as an opportunity to manipulate the primary inquirer into behaviour that is beneficial for themselves but not for the primary inquirer, rather than engage in bona fide joint ethical inquirer. Doing so would lead for the practice of joint ethical inquiry to break down, as any primary inquirer would expect others to manipulate him. Secondly,

one should not dismiss others' contribution to ethical inquiry without good reason. If doing so becomes the general practice and agents do not expect for their contributions to be taken up in good faith, this similarly leads to the breakdown of the practice of joint ethical inquiry. Finally, one should not treat others in such a way that it would be patently functional and reasonable for them to use ethical inquiry in a manipulative way to end that treatment. Doing so, would likely lead to such manipulative use and thus to the breakdown of shared ethical inquiry. Furthermore, even if it did not actually lead to such manipulation, it would lead others to reasonably suspect such manipulation.

Individual violations of these requirements do not by themselves lead to a breakdown of the practice. Hence, agents will be regularly 'tempted' to treat others in ways which would undermine shared ethical inquiry if they were more common. Hence, for shared ethical inquiry to be stable, agents must generally endorse a virtue excluding such behaviour. Thus, all maximally functional virtue-inquiry would publicly endorse it. We can call that virtue *ethical respect*. Given third-party enforcement and the very benefits of participating in shared ethical inquiry, maximally functional virtue-inquiry would also endorse it in private. Hence, ethical respect is a universal virtue.

4.3 Righteousness and third-party enforcement

In the foregoing, I have repeatedly mentioned *'enforcement mechanisms'*. These enforcement mechanisms themselves generate a tragedy of the commons. Given that agents benefit from the first-order practices maintained through such enforcement mechanisms, they benefit from the enforcement mechanisms themselves. These mechanisms contribute to effective agency by making cooperation more reliable. Yet, it can be costly to participate in such enforcement mechanisms. In *Melipona* honeybees, enforcement mechanisms are stable because the enforcing individuals derive an intrinsic benefit from the enforcement. They get nutrients from eating the eggs. This is not typically the case with humans. Active punishment has costs. In extreme cases, it involves the use of violence, and thus likely injury from the target's acts of defense. Even in less extreme cases, speaking out against injustice carries risks of *losing* reputation. Often, not cooperating with those who have manifested vices of injustice is beneficial to inclusive fitness and self-interest (on any plausible account). After all, such agents are unreliable. Sometimes, however, cooperating with them may benefit your inclusive fitness and your (apparent) interests. You may have good reason to think that they are not going to betray *you*.

Hence, agents will regularly be tempted not to participate in third-party enforcement and in order for it to be stable, there must be a virtue regulating in its favour. We can call it *righteousness*. Given the benefits to self-maintenance provided by righteousness all instances of virtue-inquiry will publicly endorse righteousness. Given this, all agents with maximally functional capacities for virtue-inquiry would also privately conclude that they ought to participate righteously in third-party enforcement.

5 Common vices and universal virtues

In this section, I will discuss two examples of practices which are incompatible with universal virtues, and which are thus immoral wherever they occur. These are chattel slavery and cruelty in war. Each is incompatible with both, virtues of moderation and virtues of justice.

5.1 Enslavement, ethical respect and patience

In this subsection, I will argue that the practice of slavery violates the universal virtues of ethical respect and patience. As such, all maximally functional instances of virtue-inquiry would converge on the judgement that slavery is impermissible. I first focus on ethical respect and then patience.

Judgements endorsing slavery as it was practiced, for example, in ancient Athens and in the ante-bellum South are incompatible with ethical respect. The enslaved cannot be expected to cooperate in genuine ethical inquiry with the enslavers or their collaborators. Hence, the slave-owning societies lacked the trust that can be produced by joint virtue-inquiry. Hence the enslavers had to be afraid of uprising by the enslaved, but also of poisoning and witchcraft (e.g. Obeah and Vodou/Voodoo/Hoodoo as seen by whites and *gens de couleur libres* in the slave-owning societies of America; cf. Savage 2007, Bryson 2013, Dwyer 2024). Such fear may be a reasonable response to potential dangers, or it may be an expression of tension or discomfort arising from the aforementioned inconsistencies. Such fear, however, harms mental self-maintenance, as well as potentially interfering with other forms of self-maintenance.

Furthermore, the violation of ethical respect made slave-owning societies unstable. Such societies are liable to experience unrest due to slave uprisings, which were known both from antiquity and from the 'New World'. Unlike with other forms of internal conflicts, such conflicts cannot be settled by relying on ethical deliberation. The enslaved who rise in insurrection have no reason to participate in

such deliberation. Furthermore, slave-owning societies are weakened in their ability to defend themselves against external enemies. Enslaved people cannot be expected to defend the community, and they cannot be armed without risking their insurrection.

In addition to universal virtues of justice, such as ethical respect, enslavement also violates universal virtues of moderation. Being an enslaver or collaborating with another's enslavement requires high levels of violence. For example, it will require very violent punishments of minor 'infractions' by the enslaved in order to ensure their subordinate status. Such violence, however, is likely to give rise to self-reinforcing feedback loops, such as those involved in wrath. Hence, slavery is contrary to the universal virtue of 'patience' (or 'gentleness'/'meekness').

This is supported by classical critiques of slavery. Frederick Douglass (2015 [1845]: 32) describes how enslavement produces vices in the enslaver, Discussing a "mistress" who had bought him and who had not owned an enslaved person before, he writes:

> But, alas! this kind heart had but a short time to remain such. The fatal poison of irresponsible power was already in her hands, and gradually commenced its infernal work. That cheerful eye, under the influence of slavery, eventually became red with rage; that voice, made all of sweet accord, changed to one of harsh and horrid discord; and that angelic face gave place to that of a demon. Thus is slavery the enemy of both the slave and the slaveholder.

Similarly, in his discussion of ancient Greek and Roman slavery, David Hume (1777) observes:

> "The little humanity, commonly observed in persons, accustomed, from their infancy, to exercise so great authority over their fellow-creatures, and to trample upon human nature, were sufficient alone to disgust us with that unbounded dominion. Nor can a more probable reason be assigned for the severe, I might say, barbarous manners of ancient times, than the practice of domestic slavery; by which every man of rank was rendered a petty tyrant, and educated amidst the flattery, submission, and low debasement of his slaves … The inhuman sports exhibited at Rome, may justly be considered too as an effect of the people's contempt for slaves, and was also a great cause of the general inhumanity of their princes and rulers.

> Who can read the accounts of the amphitheatrical entertainments without horror? Or who is surprised, that the emperors should treat that people in the same way the people treated their inferiors? One's humanity is apt to renew the barbarous wish of Caligula, that the people had but one neck: A man could almost be pleased, by a single blow, to put an end to such a race of monsters. You may thank God, says the author above cited [i.e. Plutarch] addressing himself to the Roman people, that you have a master (to wit the mild and merciful Nero) who is incapable of learning cruelty from your example. This was spoke in the beginning of his reign: But he fitted them very well afterwards; and, no doubt, was considerably improved by the sight of the barbarous objects, to which he had, from his infancy, been accustomed."

Hume's attitude to slavery is, notoriously, complicated and certainly not wholly commendable (see Watkins 2013). In the same text, he makes a passing remark that "[t]he remains which are found of domestic slavery, in the American colonies, and among some European nations, would never surely create a desire of rendering it more universal." Yet, he also personally advised his patron, the Marquess of Hertford, to acquire a plantation along with the enslaved labourers on it and facilitated the purchase. He also infamously endorsed racist views about differences in intellect between Blacks and whites. Such views were often used by others to justify slavery. Yet, this quote presents a powerful statement of the argument that slavery produces, in the enslavers, vices which make their lives worse and which make it more difficult for them to live together.

5.2 Cruelty and clemency in war

Besides slavery, another practice that appears to be common throughout much history but seems nonetheless evidently vicious to us is cruelty towards enemies in war. While warfare appears to be common throughout history, the types of violence employed in them vary drastically. In some contexts, violence is highly ritualised. It involves a clear distinction between warriors and others, with the latter not being legitimate targets of violence. Surrendering enemy combatants are captured and not killed. In other historical contexts, civilians are killed and enslaved, the women raped and/or forced into marriages with the victors, and combatants are killed whether or not they surrendered. Thus, while it may be wrong to call cruelty in warfare *typical* across human ways of life, it is very common. Nonetheless, it violates universal virtues of moderation and of justice.

Like enslavement, cruelty in war violates the virtue of patience or gentleness, i.e. the virtue regulating violence against fellow human beings. In acting cruelly towards enemies in war, combatants risk escalating a self-reinforcing feedback loop of using more and more violence. They will form a habit of using violence and see more and more situations as calling for the use of violence. This will interfere with their ability to regulate their behaviour through ethical inquiry. This can be seen, for example, by the strong and disproportional cases of PTSD in soldiers who see themselves as having committed atrocities (cf. Mohamed 2015). More generally, it can be seen from the struggles of returning soldiers to reintegrate into peaceful societies, such as in the United States after her war in Vietnam. A disproportionate amount of such soldiers became involved in organised crime or continued using violence in other, illegitimate ways. In other historical contexts, returning soldiers who were desensitised at the border of a polity to the use of violence are later used by generals against citizens in the pursuit of political aims, such as at the end of the Roman Republic. This is not necessarily to say that all such warfare was cruel. Even legitimate violence in war can have such effects. However, the effects would be exacerbated if the conduct was cruel.

Secondly, cruelty in war violates a virtue of justice. As the modern laws of war and various ethnographic examples of partly ritualised warfare show, it is possible to share some norms even with enemies in war. The existence of such norms benefits all parties in a war. Generally, the advantages to be gained by being cruel towards the enemy are less than the disadvantages from being targets of cruelty. Hence, it is a virtue of justice to uphold such norms and conventions where they are in place, and to act with a mind towards bringing them into existence where they are not. We can call this virtue, which also includes elements of a virtue of moderation, *martial clemency*.

6 Conclusion

In this chapter, I have considered the question of how we may evaluate the virtues of distant others given bio-constructivism. First, I have argued that simple deference to other communities concerning their own virtues is inappropriate due to the phenomenon of cultural pathologies. Instead, I presented counterfactual and perspective-taking ethical inquiry as ways to evaluate members of particular distant communities. Furthermore, I have argued that there are certain

universal virtues, including virtues of moderation and of justice. Such virtues are incompatible with slavery and with cruelty in war.

SEVEN: How virtue-inquiry motivates to regulate

In the preceding chapters, I had to assume that virtue-inquiry can issue in instances of 'self-control', and that such self-control constitutes a form of regulation of (the function and effects of) character traits. In order to regulate (the function and effects of) character traits, virtue-inquiry must be capable of motivating action, even if it is not involved in the usual way in which ethically praiseworthy action is motivated.

In this chapter I present two different but compatible accounts of how virtue-inquiry regulates action. I begin by providing a simple Default Account. This Default Account simply assumes that human agents have a 'desire for virtue' which interacts with ethical judgements to motivate action. This account should be acceptable to as many readers as possible. However, while it is compatible with bio-constructivism, it provides only limited evidence for it. In contrast, if successful, the Ambitious Account provides substantial additional reasons to accept bio-constructivism. The Ambitious Account is based on an 'autopoietic-enactivist' account of intentional content as presented by Evan Thompson (2007). I do not have sufficient space to fully defend the Ambitious Account, and I present it primarily as an intriguing possibility and as an avenue for further research.

1. Ethical self-control motivation and the Default Account

In order for virtue-inquiry to regulate character traits, it must be capable of motivating action. In order to see why this is the case, let us revisit the technical definition of *regulation* introduced in Chapter Four. According to that definition, a subsystem R regulates another subsystem S iff S contributes to self-maintenance by exerting a constraint F, if other subsystems presuppose that constraint (i.e. they contribute to self-maintenance if F is present) and R responds to a (partial) failure of S to exert F in such a way that makes up for some of the contributions to self-maintenance that would have otherwise occurred. For example, the human neural system regulates the heart's function to pump blood. Through pumping blood, the heart maintains a constant blood pressure. Thereby, it ensures that oxygen etc reach various organs. If the heart begins to plump less effectively, the regulatory neural subsystem responds by constricting the blood vessels. Thus, it ensures that at least some oxygen still reaches the organs. Thus, it may be more in line with ordinary usage to say

that the neural subsystem regulates *the effects* or *the function* of the heart. However, for brevity's sake, I also use the phrase, *regulates the heart* (or: *regulates character traits*).

One implication of this concerns the use of expressions like *well-regulated* and *dysregulated* character traits as well as *well-regulated* and *dysregulated* emotions, as they are used for example in psychology. According to my technical terminology, it makes most sense to say that character traits and the emotions expressing them are well-regulated if their effects are the effects that are regulated for. This may be the case because there is an active and effective regulatory subsystem. However, it is at least as likely that the effects are this way because the character traits are such that they produce these effects without the regulatory subsystem intervening. Thus, often, a well-regulated character is a little regulated character: a character that makes regulation unnecessary.

As we have seen in Chapters Four and Five, regulatory subsystems in part *determine* the functions of regulat*ed* subsystems. It is the presence of a regulatory subsystem that distinguishes incidental benefits to self-maintenance from full-fledged functions. Ethical inquiry, I have argued, follows its own peculiar logic precisely because a central part of it, virtue-inquiry, plays an important role as a regulatory subsystem in *determining* the functions that it simultaneously *describes*. Thus, in order for the account of ethical inquiry given in Chapter Five to be successful, virtue-inquiry must regulate character traits, i.e. it must be capable of counteracting the effects of character traits. Character traits contribute to self-maintenance through producing actions. Hence, virtue-inquiry itself must also be able to produce behaviour, i.e. to *motivate*. Through motivating action in spite of countervailing character traits, virtue-inquiry also contributes to the forming of habits and to changing the character traits. This effect is of immense practical importance to human agency, but it is ancillary to my technical notion of *regulating character traits*.

For illustration, consider an agent who judges that they ought to be less timid and more daring because that would make them more courageous. Holding on to this judgement, they find themselves in a situation where they experience some inner conflict. They see a child drowning in a somewhat dangerous sea. They judge that they would very likely be able to save the drowning child and themselves if they jumped in. However, they are also fearful of drowning and their fear is holding them back from jumping in. Without their ethical judgement, they would have remained on land. However, their ethical judgement somehow intervenes, and they save the child. In this scenario,

virtue-inquiry has regulated (the effects of) the agent's character trait. In particular, it has regulated (the effect of) their attitude to danger (i.e. the character trait that is courage if virtuous and cowardice or recklessness if vicious). In order to do so, virtue-inquiry had to be capable of playing a pivotal role in motivating the agent's action, viz. jumping into the water to save the child.

We may call this form of motivation, *ethical self-control motivation*. It should be distinguished from what has been called *ethical motivation* more generally. This is the form of motivation involved in actions that are generally called 'ethical', e.g. acts of altruism and selfless rule-following. According to virtue-ethical accounts, including bio-constructivism, such actions may often not involve ethical judgements at all.

In principle, it may be that an ethical judgement *alone* motivates an action like saving the child. The Default and the Ambitious Account will both be compatible with this. However, it may be more likely that ethical judgements determine action only if there is a motivational conflict anyway. If an agent does have *any* motivation to save the child independently of their judgement that doing so would be courageous, they may be too wicked for that judgement to make them jump into the water. It is sufficient in order for virtue-inquiry to be a regulatory subsystem if ethical judgements can reverse the outcome in cases of tight internal motivational conflicts. This still requires ethical judgements to be capable of motivating action.

It may seem puzzling how ethical judgements could have the power to motivate me at all. After all, according to bio-constructivism ethical judgements are merely a special kind of judgements about organisational functions, namely judgements about the organisational functions of character traits. To be precise, they are primarily judgements about the functions of the judging agent's own character traits, and derivatively about those of others'.

This problem is exacerbated due to a realist commitment bio-constructivism shares with neo-Aristotelian naturalism and other forms of function-based naturalism: Ethical judgements are distinguished from other judgements by their *content* and their *subject matter*. Insofar as they are some special mental state which is distinct in its powers from other judgements, this is to be explained by their subject matter.

However, seemingly, agents may make judgements about the organisational function of their character traits without being motivated by them. Imagine an agent familiar with the organisational account of function, but in disagreement with bio-constructivism and the organisational account of virtue, who could make a judgement which they would express as 'If I was more daring, my character trait of courage would actualise its organisational function; as things are, it is malfunctioning.' Such an agent may well not be motivated in any way to be or become more daring. It is no more necessary that he should be motivated in line with that judgement than an agent must be motivated to be(come) jealous who judges 'If I was more jealous, my character trait of sexual suspicion would realise its *selected-effect* function; currently it is dysfunctional according to selected-effect accounts'.

1.1 The Default Account

A first-pass response to this problem is to simply assume as a brute fact that virtue-inquiry can motivate action. If one assumes belief/desire-psychology, this can be put by saying that agents typically have a *desire for virtue*: a desire (or other conative attitudes) to behave in those ways one determines to be virtuous in virtue-inquiry, whatever they may be. This can be described as a desire to virtuous '*de dicto*' insofar as it goes beyond assuming that agents have *de re* desires to behave in those ways which are in fact determined to be virtuous by virtue-inquiry. However, it does not require the agents to have an articulable concept of *virtue*. Given this desire, virtue-inquiry could regulate action.

1.1.1 Moral fetishism and self-effacement

One objection this is often raised to such accounts is that it is unacceptably self-centered and 'fetishistic' (Smith 1994). Morally exemplary agents do not act out of a desire to be virtuous. For example, if a morally exemplary agent jumps into the cold water to save a drowning child, they do so because they desire for the child to live. They do not do so because they desire for themselves to be courageous. Acting on a desire to be virtuous makes a fetish out of virtue. However, bio-constructivism may agree with that assessment. The maximally well-functioning agent would not be motivated by their virtue-inquiry. Instead they would be motivated in line with their virtues. The virtue of charity involves treating the considerations that the child is drowning and that I can save them as a *pro tanto* reason to attempt doing so; the virtue of prudent caution involves treating the consideration that doing so would be somewhat dangerous as a lesser reason not to do so. Thus, the

virtuous agent will be motivated simply by the considerations that the child is drowning. Possibly, this can be explained by saying that they have a desire not to let the child drown. Only the less-than-fully-virtuous agent is motivated by their virtue-inquiry. It must still be the case for the virtuous agent that virtue-inquiry regulates in favour of their virtuous character traits. If they *were* to fail to act virtuously, virtue-inquiry *would* produce a counteracting motivational push. Thus, they still must have a desire for virtue. However, there is nothing 'fetishistic' about wanting to be (more) virtuous. It is only fetishistic for such a desire to be stronger and more important than the desires constitutive of particular virtues.

The worry about fetishism relates to worries about *self-effacement*. This worry was first introduced by Stocker (1976). Stocker asks us to imagine a committed utilitarian who visits her best friend in hospital. On being asked why she visited, she explains that she realised that because of her special relationship with her best friend, this was the optimal use of her time to maximise overall happiness. Her friend would be rightly upset, and might say: 'If that's your motive you don't need to turn up in the first place.' They might well react similarly if their friend was a deontologist whose motive was to discharge their impartial duty. This seemingly shows that a good or praiseworthy agent could not be motivated by the considerations which justify her actions according to these two theories (cf. Pettigrove 2011). Thus, knowledge of the theories would seemingly undermine the ability to perform good actions, and the theories may describe their own suppression. Hurka (2001) and Keller (2007) argue that the same objection applies to virtue ethics. Hurka attacks forms of virtue ethics according to which an action is justified because it promotes flourishing. Keller attacks forms according to which an action is justified because it is what a virtuous agent would do.

Whatever the merit of such arguments against virtue ethics (see Annas 2008 and Pettigrove 2011 for responses), it does not affect bio-constructivism under the Default (or, as will see, the Ambitious) Account. Fully virtuous agents according to bio-constructivism are not motivated by the desire for virtue at all. Rather, they are motivated by the considerations which they treat as reasons. The virtuous agent visiting her friend would truthfully reply: 'I am visiting you because you are my friend and you need me'. The less-than-fully-virtuous agent would have to reply 'Well, I want to support you because you are my friend. But to be fully honest, I wasn't sure about whether to come or not until I realised that in order to be virtuous, I should visit you.' The friend in hospital would be

rightfully disappointed with this response; however that is what bio-constructivism predicts. The second visitor is less than fully virtuous.

Knowing about bio-constructivism would not transform a virtuous agent into the less-than-fully-virtuous agent, or motivate them to try doing so. Hence, bio-constructivism provides no reason. Furthermore, despite the connection between self-maintenance and virtue, even the less-than-fully-virtuous agent is not motivated by a selfish desire *de dicto* for self-maintenance. Rather, they are motivated by a desire for virtue. Virtue merely happens to be *de re* identical with a certain kind of function founded in self-maintenance. Self-effacement is still a relevant worry for ethical theories constructed under bio-constructivism. Given the diachronic consistency constraint, virtue-inquiry may still involve constructing ethical theories. It counts against such theories if they are self-effacing. This is due to the shareability constraints. If an ethical theory is self-effacing and believing in it interferes with being a good agent, it is in my interest not to share it with others. Thus, it would likely not be endorsed by ethical inquiry actualising its function.

1.1.2 Frege puzzles and modes of presentation

The Default Account can also avoid the problem that a judgement expressible as 'if I was more daring, my character trait of courage would actualise its organisational function; as things are, it is malfunctioning' may fail to motivate. Given the Default Account, this appears as a version of so-called *Frege puzzles*. Here is a more conventional Frege puzzle (cf. Mann and Pain 2022a, Heck 2012). Elizabeth reads in the morning paper that George Eliot has died. She is mildly interested, since she knows some of the novelist's work. Later that day, she learns that her sensitive and intellectual cousin Mary Ann Evans has died. She liked Mary Ann, and is affected by the death. She begins making preparations to attend the funeral. Unbeknownst to her, George Eliot and Mary Ann Evans are the same person. Thus, apparently the belief that 'Mary Ann Evans died' combined with a desire to mourn her cousin's death appropriately produced action. Yet, the belief that George Eliot died didn't. This is the case even though both beliefs had the same truth-conditions.

One common response to such puzzles (and Frege's own) is to posit *modes of presentation* (or *'senses'*). Modes of presentations are aspects of meaning or content other than truth-conditions. Thus, *George Eliot* and *Mary Ann Evans* are associated with different modes of presentation in Elizabeth's mind.

Just like beliefs, the contents of desires also have modes of presentation. Elizabeth's desire to mourn George Eliot/Mary Ann Evans features that mode of presentation linked with *Mary Ann Evans*, but not that linked with *George Eliot*. Similarly, the desire for virtue plausibly features the mode of presentation linked with *virtue* but not that linked with *organisational-functional character trait*.

1.1.3 But do humans have a desire for virtue?

There are some considerations to suggest that such a desire to be virtuous is common among human beings and that in it virtue is presented in such a way that allows the desire to combine with the judgements of virtue-inquiry. Apparently, such a desire is necessary for virtue-inquiry to regulate character traits and the actions they produce. However, virtue-inquiry regulating actions produces effective agency and thereby contributes to the self-maintenance of the lineage directly as well as indirectly through the self-maintenance of the individual organism. Thus, it may well be selected for. As a selected-for trait, we can expect such a desire to be common among human beings. Furthermore, given the usefulness of such a desire both for the individual itself and for all members of his or her community, it is likely that such a desire will commonly be further inculcated into children during their upbringing, for example through reward and punishment and through the presentation of admired exemplars. Some readers may be happy to accept it as an assumption that human agents typically have such a desire for virtue. Velleman (2014) for example, argues that agents have a similar desire to be *intelligible*, which motivates them to act in line with their ethical judgements. These readers may disregard the rest of this chapter.

1.1.4 The limitations of the Default Account

Some readers, however, may find the account so far unsatisfying. The desire for virtue is contingent. For all that has been said so far, there could be human beings without it. Furthermore, such human beings could not be ethically criticised for lacking it. Their motivational system would not be malfunctioning for lacking it. Rather, such agents would not be proper targets of ethical evaluation at all. Insofar as their character traits have regulated functions, these functions would not be determined by virtue-inquiry. Thus, others could not rely on ethical inquiry, with its distinctive logic, to inquire whether their traits are functional or not.

Secondly, since the desire for virtue is simply assumed, the account provides only limited evidence for the identification of organisational function with virtue. There are two primary points of evidence: (a) the ability of the organisational account to capture judgements about biological functions as well as competing accounts, and in some cases better, such that it is a legitimate account of biological function, together with the similarity of the words and concepts used to talk about virtue and biological function; (b) the ability of bio-constructivism to explain the logic of ethical inquiry. However, bio-constructivism can only explain the logic of ethical inquiry *if* we assume that agents have a desire for virtue which is linked to virtue-inquiry. Furthermore, we must assume that the object of this desire is identical with the organisational function of character traits. This provides some evidence for the organisational account. After all, there are many other presuppositions that could have been made which would not have led to an adequate account of ethical inquiry. This would be the case, for example, if we presupposed that virtue was identical with *selected-effect* functions or propensity to advance fitness (see Chapter Two). Nonetheless, we cannot exclude that other presuppositions would produce an equally adequate account of ethical inquiry.

We would have stronger evidence for bio-constructivism if we could show independently that human agents are typically motivated to promote the organisational functions of their character traits (under some mode of presentation). If so, we could identify this motivational disposition with the 'desire for virtue'. In doing so, we would have an argument to replace the somewhat arbitrary assumption that all agents have such a desire. We would further have an argument that even agents who appear to lack the desire for virtue have it in a rudimentary form, and are thus appropriate targets for ethical evaluation according to bio-constructivism.

Finally, we would not have to presuppose that the desire for virtue concerns the organisational function of character traits (under a different mode of presentation). Rather, we'd have an independent argument that human agents are motivated to promote the organisational functions of their character traits (under some mode of presentation). Given this desire, they would be motivated to inquire into these organisational functions. Due to the arguments of Chapter Five, this inquiry would have a logic very much like that of ethical inquiry as we ordinarily know it. Thus, we would have stronger evidence to identify virtue with the organisational function of character traits, this inquiry into the organisational function of character traits with virtue-inquiry and the desire to

promote the organisational functions of one's character traits with the desire for virtue. In the following, I will provide just such an argument.

1.2 Belief/desire-psychology

So far, in this chapter, I have implicitly assumed *belief/desire-psychology*, and spoken in ways that suggest ***mechanistic*** *belief/desire-psychology* in particular. According to *belief/desire-psychology*, there are two fundamental kinds of contentful mental attitudes: *Cognitive* and *conative* attitudes. Cognitive attitudes concern what is the case. They are beliefs and judgements as well as perceptions, guesses, hunches, etc. If the content of a cognitive attitude is incorrect, the attitude or its possessor is in a certain sense at fault, and she ought to give it up. In contrast, conative attitudes can be described as concerning what ought or should be the case (as far as their possessor is concerned). They are, for example, desires, wants, wishes, appetites and preferences.

According to belief/desire-psychology, every action can be explained by reference to a conative and a cognitive attitude of the agent's. The conative attitude describes some goal in the action. The cognitive attitude has the content that an action of the type that is being performed is an adequate means to bring about that end. Often such explanations may be enthymematic in that the cognitive attitude appealed to is one that *implies* that the action is an adequate means for the goal. For example, I might explain my roommate's going to the fridge by saying that she wants to drink milk and that she believes that there is milk in the fridge.

According to *mechanistic* belief/desire-psychology, cognitive and conative states are distinct psychological entities which interact in causal ways to produce action. One can accept belief/desire-psychology without accepting mechanistic belief/desire-psychology. Instead one may for example accept an interpretivist form of belief/desire-psychology: any agent can appropriately be ascribed both conative and cognitive states based on any of their actions, and in order to explain an action in this way, both a cognitive and a conative state must always be ascribed. However, the different ascriptions do not necessarily correspond to different causally efficacious entities. Similarly, dispositionalist forms of belief/desire-psychology might identify cognitive and conative states each with different dispositions. Any action involves both desire-like and belief-like dispositions. Yet these dispositions might each be explained by a shared foundational structure. Compare how the same

object may be both fragile and water-soluble in virtue of the same crystalline microstructure, yet fragility and water-solubility are different dispositions.

The Default Account assumes belief/desire-psychology, and is compatible with both mechanistic and non-mechanistic forms of it. The Ambitious Account will deny mechanistic belief/desire-psychology (at least as an explanation of *all* action), but be compatible with non-mechanistic forms of belief/desire-psychology.

2. The Ambitious Account

In this section, I describe an Ambitious Account of ethical self-control motivation. This account is more speculative than the Default Account. It draws on bio-constructivism's account of the (truth-conditional) content of virtue-judgements to explain their motivational force. Therefore, if successful, it provides strong additional support for bio-constructivism. According to the Ambitious Account, intentional states with a truth-conditional content that pertains to self-maintenance *ceteris paribus* motivate an agent to act.

In this section, I first introduce the Ambitious Account as such (Section 2.1). Then, I introduce autopoetic enactivism and consider some objections to it (2.2). Finally, I consider an important objection to the Ambitious Account: if virtue-judgements motivate in virtue of their truth-conditional content, how can some judgements have the same truth conditions without motivating (2.3)?

2.1 Introducing the Ambitious Account

According to the Ambitious Account, some 'cognitive' states have a motivational force in virtue of their *truth-conditional content*. Their truth-conditional content consists in the conditions under which they are true. These are identical across modes of presentation. Specifically, cognitive states motivate action in virtue of their truth-conditional content if that truth-conditional content concerns the agent's own self-maintenance. This includes cases in which it concerns organisational functions. Specifically, if a representation with non-derived intentionality has the truth-conditional content that a certain behaviour would significantly contribute to self-maintenance or prevent significant harm to

self-maintenance, then it *ceteris paribus* motivates the organism whose representation it is to behave in such a way.

The *ceteris paribus* provision accounts for the fact that not *every* state with such a truth-conditional content in fact motivates. However, if a state with such a truth-conditional content doesn't motivate, this is due to the presence of peculiar, precluding features of that state. It is not due to the *absence* of some further, desire-like states that are present when it motivates.

One example of a representation that motivates solely in virtue of its truth-conditional content is the perceptual representation of a dangerous ball approaching me. Given a plausible reading of *dangerous*, this representation has the truth-conditional content that there is a ball approaching such that if I do not avoid it, it will cause serious damage to my bodily integrity and thus to my ability to self-maintain. In virtue of this truth-conditional content, it will motivate me to dodge the ball. Other examples include the judgments that some food is poisonous and others wholesome and nutritious. They will motivate me to eat the latter and avoid the former.

According to bio-constructivism the truth-conditional content of ethical judgments similarly concerns self-maintenance (on various levels). If I judge that saving a drowning child would be courageous for me, the truth-conditional content of my judgement is that doing so would contribute to self-maintenance. (In this case, the relevant self-maintenance is that of an extended or quasi-lineage, see Chapter Four.)

One might object that in ethical judgement an action is presented merely as being the product of a trait that typically contributes to self-maintenance, rather than as contributing to self-maintenance itself. Thus, the case differs from paradigmatic cases like the dangerous ball. However, even in cases like the flying ball, there is some uncertainty in my awareness. The ball might miss me. I might hurt myself in my evasive manoeuvres. Being hit by the ball might fix my chronic headaches. Thus, both avoiding the ball and saving the drowning child are presented as contributing to self-maintenance, though they are presented as such with less than maximal certainty.

One may also worry that it is unrealistic that representations whose truth-conditional content concerns the contribution our actions could make to the self-maintenance of extended or

quasi-lineages *ceteris paribus* motivate. Surely, there is nothing unusual about a human agent judging that they could help another through an action, but not being motivated in any way to perform that action? However, firstly, as we will see in Section 2.1, a plausible account of the motivational force of judgements concerning individual organismic self-maintenance naturally extends to such cases.

Secondly, the fact that such judgements 'usually' do not motivate can be accommodated by the Ambitious Account. As I will describe in 2.3, judgements whose truth-conditional content concerns self-maintenance may fail to motivate if they are made in a derivative way that allows the agent to make them without imaginatively engaging with them. It is possible that it is very easy to make judgments that concern the self-maintenance of extended and quasi-lineages in such a way. Especially, it may be easy to present the possibility to contribute to the self-maintenance of quasi-lineages in a third-personal way.

One may also think that it is simply implausible that the mere truth-conditional content of a state could explain its motivational force. After all, representations can have the same truth-conditional content --- be true in all of the same possible worlds --- and yet be completely different. It may turn out that a belief that, in an intuitive sense, has 'nothing to do' with self-maintenance is true in all and only the same possible worlds as belief the agent would describe in terms related to self-maintenance (such as *danger*, *disease*, *harm*). Thus, they have the same truth-conditional content. But why should the former belief have any motivational force?

However, in order to have the same truth-conditional content, two representational states must share the same truth-value over a very wide set of possible worlds. This includes where the agent's chemical structure is radically different (or has been radically changed through stepwise adjustments). Hence, for example, the belief that this is water and that this is a relatively chemically pure liquid that will quench (the need indicated by) my thirst and have little other effect on my body have different truth-conditions. My body could be constituted radically differently. Thus, it could be such that my thirst was not quenched by water. If beliefs have truth-conditions that track contributions to self-maintenance across all such possible worlds, then, I suggest, they will *ceteris paribus* have motivational force simply in virtue of having these truth-conditions.

In some respects, the Ambitious Account is similar to the position defended by McDowell (1998). McDowell also denies mechanistic belief/desire-psychology and holds that some judgements motivate simply by virtue of the content they have as cognitive states. (Although he is less clear on whether they motivation solely by virtue of their *truth-conditional* content, or if modes of presentations are relevant.) However, McDowell holds that accounting for the content of such beliefs requires a 'liberal' notion of nature and 'natural' facts. The truth conditions cannot be describable by natural science. Enoch (2011) goes even further. He holds that it requires admitting *non*-natural facts. In contrast, the Ambitious Account suggests that a belief could have this dual character even though its truth-conditions are in principle fully describable by natural science. I do not have space to fully defend this suggestion against most criticisms McDowell or Enoch could raise against it. Rather, my aim is to make the account plausible. I mention McDowell's and Enoch's positions not to disprove them but to throw my suggestion into relief.

2.2 Autopoetic enactivism

The connection between truth-conditional content pertaining to self-maintenance and motivation can be explained by an account of intentionality due to Evan Thompson (2007). This account has been called "autopoetic enactivism" (see e.g. Hutto and Myin 2013). Drawing on Weber and Varela (2002), Thompson argues that life consists in a particular form of self-maintenance and -creation, or *autopoiesis* and is closely connected to intentionality and cognition. I will also use the terms *intelligence* and *intelligent behaviour*. This framework is a precursor to the organisational approach to organisms and their functions.

In order to maintain their organisation, living beings need to manipulate their surroundings or their relation to them. For example, they need to acquire nourishment. Consider the locomotion of *E. coli* bacteria. In order to maintain themselves, they must be in a sugar-rich milieu. Given this, *E. coli* move randomly until they detect an upwards gradient of sugar concentration in their direction of movement. Then, they begin to move consistently in the same direction. They do so until they detect the sugar gradient falling, in which case they resume random movement. As Thompson suggests, this is a most rudimentary example of "cognition" or 'proto-cognition' and intelligent behaviour in the broadest sense. At the same time, it is a rudimentary example of (proto-)*intentionality*. Intentionality is often described as the property of mental states to be about

something. However, Thompson takes it that the more fundamental phenomenon is the world being *disclosed* to a living being, there being a world *for* that being. The bacterium is able to respond intelligently to the sugar-gradient in its surroundings. The sugar-gradient is there *for* the bacterium in the most elementary sense.

This basic form of (proto-)intentionality is valenced. A rising sugar-gradient is positively valenced: it 'calls on' the bacterium to continue in its current movement. A falling (or steady) sugar-gradient is negatively valenced. It calls on the bacterium to move randomly until a rising sugar-gradient is disclosed to it. Human cognition, while being vastly more complex and sophisticated, rests on the same basis. A world is disclosed to us which has a practical significance for us. Thus, intentionality is essentiality connected to valence. Without valence, disclosures would not be relevant to a living being's intelligent behaviour. Thus, the world would not really be disclosed to it. Furthermore, valences are intrinsically connected to self-maintenance. Positively valenced disclosures must typically be of features that contribute to self-maintenance. Negatively valenced ones must typically be of features that inhibit self-maintenance. Of course, there may be exceptions to this typical correlation. However, if it was not typical, then the system's behaviour would not be intelligent and the system would not exhibit intentionality. Nothing would be disclosed to it.

Using the terminology of contemporary neo-Aristotelians, we can say that a notion of a life-form is incoherent to whom some features that contribute to its self-maintenance are characteristically disclosed with a negative valence or features that inhibit self-maintenance are characteristically disclosed with a positive valence. Such a life-form could only survive due to chance, but life-forms must be able to survive through their own characteristic behaviour. Hence, such a life-form could not exist.

Evan Thompson (2007) makes no explicit reference to truth-conditional or other contents. Instead, he speaks of the world being "presented" to the creature, as well as about "meaning" and "significance" emerging. However, it is plausible that given his account, we can ascribe states with truth- or correctness conditions to the bacterium. This is suggested by the idea that the world is 'presented' to the bacterium. The 'world' presumably must supply an independent standard against which the cognitive states of the agent can be evaluated.

Furthermore, Thompson hopes that such proto-intentionality may help to explain the forms of intentionality found in humans and other higher animals. Their states, however, have truth-conditional content. Thus, the project of explaining sophisticated intentionality on the basis of such foundational (proto-)intentionality is more plausible if (proto-)cognition already features truth- or correctness-conditions. These correctness-conditions can plausibly be identified with the conditions under which the relevant (sub)systems of the bacterium realise their functions. Thus, the correctness-conditions of the bacterium's state if it is swimming (say) northwards is their being an increasing gradient of nutrients for the bacterium towards North, i.e. an increasing gradient of substances necessary for the bacterium's self-maintenance.

This does not assume that there is a separable state or entity within the bacterium which represents this state of affairs. Rather, a state with this truth-conditional content may be ascribed to the bacterium based on its interactions with its environment. We may presume that similar behaviour may be occasioned in the bacterium by some non-sugar molecules that are not nutritious for the bacterium (similar to how artificial sweeteners act for human beings). If the bacterium is responding to such substances in its environment, then it is in an incorrect (or false) cognitive state.

Similar accounts, which ground intentionality in organismal self-maintenance and self-regulation have been defended by Hooker (2009, 2024), Mann and Pain (2022b), Rama (2023, 2024), Sachs (2019), González de Prado Salas (2018), Schroeder (2004) and Bickhard (2000, 2004, 2009, 2024). Citing van Duijn, Keijser and Franken (2006), Peter Godfrey-Smith (2016a, see also 2016b,c, 2022, 2024) similarly defends an account of 'minimal cognition', according to which cognition and "metabolism", or self-maintenance are closely linked.

2.2.1 Connecting bio-constructivism and autopoetic enactivism

According to bio-constructivism, virtues are functional, i.e. traits that typically contribute to self-maintenance. For an action to be virtuous, i.e. expressive of virtue, thus suggests that it contributes to self-maintenance. Hence, it is unsurprising that, if an agent has a way of being aware of a potential virtuous course of action, that awareness would be positively valenced. Thus, it is capable of motivating action without an additional conative state.

Connecting Thompson's autopoetic enactivism with bio-constructivism in this way requires some adjustments and reinterpretations of the former. Thompson and Varela only consider one level of self-maintenance, that of the individual organism. In contrast, bio-constructivism assumes that there are multiple levels of self-maintenance. However, Varela/Thompson's approach can be combined with this assumption. In fact, it must be in order to be plausible. We can imagine a simple creature, similar to *E. coli*, which rather than a nutrient-rich environment seeks out an environment that is beneficial to reproduction in particular. It would be highly arbitrary to deny that such a creature engages in intelligent behaviour whilst maintaining that *E. coli* does. However, the seeking behaviour of our simple creature only contributes to self-maintenance on the level of the lineage. Yet, this can be described as follows: The *organisation* of an organism is a way for matter and energy to be arranged into a self-maintaining system. Living beings generally share their organisation with their offspring, *modulo* some relatively minor changes. The (instance of the) organisation of our simple creature self-maintains through reproduction, and hence through seeking out the environment conducive to reproduction. Hence, for this organisation, that environment is presented with a positive valence. Similar points apply to extended lineages and quasi-lineages, as the organisation self-maintains through maintaining other, similarly organised organisms. Finally, mental systems can be seen as having their own organisation, which is to some degree independent, and to which certain considerations can appear with a positive valence.

In human beings, these different levels of self-maintenance are integrated through a common regulatory system, i.e. virtue-inquiry. Traits are only virtuous if they contribute to self-maintenance *and they are regulated for by virtue-inquiry*. One might wonder why we should have a positively valenced way of being aware of such traits rather than simply of traits that contribute to self-maintenance. However, as we have seen, with humans' faculty of reason, many potential traits could contribute to self-maintenance on some level and in some way. Many of these traits are incompatible with each other. If human beings had positively valenced modes of awareness of all such contributions, or of a representative subclass of them, then their conduct in this regard would be inconsistent and self-undermining. It would not contribute to self-maintenance and it would not be recognisable as intelligent behaviour. Thus, only a subset of traits contributing to self-maintenance can be disclosed with a positive valence. In order for the resulting behaviour to be coherent, these must be traits that are reinforced by other aspects of the agent's behaviour, i.e. traits that are regulated for. Hence, the only traits of reason that are typically disclosed with a positive valence or those that are regulated for.

Thus, autopoetic enactivism provides an account of ethical motivation that does not rely on mechanistic belief/desire-psychology. Ethical judgements of an agent motivate because they concern the organisational function of their own traits and thus their own self-maintenance. Judgements (and other intentional representations) concerning one's own self-maintenance are inherently motivating. This is a fundamental fact about the nature of intentionality and intelligence. Such motivation is in an important sense more basic than motivation as per mechanistic belief/desire-psychology. This account is compatible with non-mechanistic belief/desire-psychology. It may be that all intentional agents capable of ethical inquiry and virtue-inquiry can rightly be said to have a desire for virtue. However, this desire is not a separate psychological entity which they might lack. Under dispositional forms of belief/desire-psychology, the desire for virtue is a disposition ethical agents have simply in virtue of their intelligence and intentionality. According to versions of interpretivist belief/desire-psychology, any living ethical agent would necessarily behave such that one could ascribe to them a (defeasible) desire for virtue. For convenience, I will continue to speak of a 'desire for virtue' in the remainder of this chapter. This should be taken in a non-mechanistic way. Furthermore, such talk can be replaced with talk about 'being motivated by ethical judgements' by readers who disagree with all forms of belief/desire-psychology (cf. Nagel 1970).

2.2.1 Objections to autopoetic enactivism

In this section, subsection, I will consider a number of objections to the form of autopoetic enactivism the Ambitious Account relies on. On the one hand, one may be worried that autopoetic enactivism cannot account for human beings' varied motivations and valenced intentional states. On the other hand, it may seem implausible that such simple creatures like *E. coli* have any contentful intentional states.

First of all, it might be objected that it is implausible that all instances of positive valence in the human case can be connected to self-maintenance. One might think this irrelevant. After all, strictly speaking, the Ambitious Account only requires that a truth-conditional content concerned with self-maintenance is *ceteris paribus* **sufficient** for a representation to motivate. It does not require it to be necessary. Thus, one might think that the Ambitious Account can easily admit that some representations are positively valenced without being connected to self-maintenance.

Yet, autopoetic enactivism aims to be an account of intelligent action and intentionality. If there could be representations with underived intentionality that have a positive valence and motivate behaviour but do not relate to self-maintenance, then this raises the question why the positive valences associated with representations that *do* concern self-maintenance could not be of the same type. Thus, the motivational force would become, in principle, independent of the truth-conditional content. Autopoetic enactivism would be replaced with a version of mechanistic belief/desire-psychology. It could not give a distinctive account of ethical self-control motivation based on bio-constructivism.

However, it is plausible that all valenced representation can be connected to self-maintenance. I will consider the positive valence associated with humour as an example. Pre-theoretically, humour seems about as far from self-maintenance as any form of positive experience. Firstly, humour has an important social dimension. If I find something funny, often I typically quickly think of sharing it with others. Doing so will contribute to my reputation with peers and thus contribute to my self-maintenance. Thus, in finding something funny, the humorous object may be represented with a positive valence partly because it is an opportunity for bonding and building reputations.

Secondly (and more generalisably), this can be explained by the dual nature of the human capacity for pleasure. I can only sketch this dual nature here without defending my suggestion. The capacity for pleasure corresponds, roughly, to what is called the 'reward system' in cognitive neuro-science. On the one hand, pleasure *indicates* contributions to self-maintenance. Episodes of pleasure have truth-conditional content related to self-maintenance. On the other hand, pleasure *contributes* to self-maintenance on the mental level. It indicates that things are going well. Thus, if an agent is 'starved for pleasure', this indicates that things are not going well. It would be adaptive for them to undergo a radical re-organisation of their behavioural dispositions. Hence, the lack of pleasurable experience is apt to cause a mental breakdown: a failure of self-maintenance of the mental organisation. Hence, on a lower level, my reward system represents the humorous object as contributing to self-maintenance on an organismic level. For example, following a popular evolutionary account of humour and laughter (Gervais & Wilson 2005), the form of pleasure which is distinctive of humour and often accompanied by laughter may indicate that a situation assumed to be dangerous or threatening to one's reputation is actually harmless. Think of how we laugh when it

turns out that the 'dangerous animal' that was scaring us was our lapdog. Most occasions for humour share a similar structure of incongruity that is first set up and then resolved. On this lower level, the truth-conditional content of representations associated with humour may often be false or inappropriate. We can experience the pleasure associated with humour even if we had not previously experienced the situation as dangerous. On a higher level, we may judge something to be funny. This judgement is positively valenced due to its truth-conditional content including the humorous object's contributing to mental self-maintenance by eliciting the distinctive pleasure of humour.

Of course such a positive valence does not always lead to action. In fact, a representation that is positively valenced may even provide a reason not to act in the way that is presented with a positive valence. For example, the judgement that a certain action would be funny provides me with a reason *not* to perform it at a solemn occasion. Even the judgement that some food is wholesome and nutritious may provide me with a reason not to eat it if I am a method actor preparing for a role that requires an emaciated figure. However, in both of these cases, there are inferences from the original judgements which have a negative valence. From the judgement that an action would be funny, I can infer that it is *inappropriate* at the solemn occasion. From the judgement that the food is wholesome and nutritious, I can infer that would infer with my goal of getting an emaciated figure and thus with my professional goals as an actor. Thus, these (implicit or explicit) conflicting inferences motivate the action I ultimately take. The original judgements still have motivational force. We would not be surprised if the method actor 'gave in' and ate the wholesome and nutritious meal in the way we would be surprised if someone ate a meal they knew to be poisonous. We are also not *puzzled* if someone tells a funny but inappropriate joke at, say, a funeral.

Due to their connection to virtue-inquiry, judgements of first-order ethical inquiry similarly have truth-conditional contents involving self-maintenance. The judgment 'I ought to save the drowning child' is true iff it would actualise a virtue (the regulated-for function of a character trait), i.e. iff saving the drawing child expresses a character trait contributing to self-maintenance in a regulated-for way. Plausibly, there may also be practical conclusions in ethical inquiry that do not have truth-conditions strictly speaking. They may be expressed as self-directed imperatives, i.e. 'let's go save the drowning child!' While not having truth-conditions strictly speaking, the autopoetic enactivist can assume that such states have *correctness-* or *appropriateness-conditions* which play a similar role. The judgement is only appropriate if saving the drowning child would actualise a virtue.

One may also worry that autopoetic enactivist could not explain all human motivation because autopoetic-enactivist motivation is *selfish*, but some human motivation --- first and foremost a lot of ethically praiseworthy motivation --- is unselfish. This is closely related to the self-effacement worry discussed in relation to the Default Account in 1.1. The latter is the worry that knowing about bio-constructivism and using it to justify my actions would make me selfish. The former is the worry that autopoetic enactivism means I am *already* selfish, even without knowing about bio-constructivism.

Neither worry is apt. The agent's motives relate to the *modes of presentation* of the states motivating them, not to their bare truth-conditional content. Hence, the agent saving the drowning child is motivated by the consideration that the child is drowning and she can save it. The connection of that consideration to self-maintenance explains its motivational force, but it does not itself motivate her. Even the less-than-fully-virtuous agent is motivated by the thought that saving the child would be *virtuous*. Again, the connection between virtue and self-maintenance explains that motivational force; it does not itself motivate. Furthermore, bio-constructivism and the organisational account of virtue do not constitute a way of justifying ethical judgements. Rather, ethical judgments are justified within ethical inquiry according to its own, autonomous logic. These justifications are unlikely to appeal to self-maintenance. Rather, bio-constructivism quietens anxieties about the place of ethical inquiry as a search for truth in a world explainable by natural science.

On the other hand, it may be objected that comparatively simple organisms like *E. coli* cannot have any representational states. One might especially worry that if bacteria and other unicellular organisms are capable of intelligence and intentionality, so must be plants. Yet, so the objection goes, plants evidently are *not* intelligent.

However, firstly it is plausible to describe plant behaviour as intelligent. Through their growth patterns, they *forage* for nutrients. Such behaviour is functionally equivalent to animals foraging for food (cf. Hiernaux 2023). Secondly, it is also tenable to say that unicellular organisms exhibit more intentionality than plants. Plants have much less of an 'identity' or a 'self/other boundary' compared to both animals and single-cell organisms. Their life processes do not distinguish between other parts of the same organism and different organisms of the same species. This is why grafting is

possible, and why plants may be cloned simply by taking a cutting. It also partly explains why most plants have less defined 'body-plans' than animals. A tree may have any number of branches without dysfunction, but the number of limbs of a mammal is fixed. Hence, it would be tenable to say that single-cell organisms exhibit (proto-)intentionality, but that vegetal organisms do not. On such a view, plants are extremely closely cooperating colonies of cells which exhibit (proto-)intentionality individually, but not collectively.

Arnellos and Moreno (2022) maintain that 'cognition' depends on the presence of a nervous system, a (sub)system that is decoupled from systems that contribute to self-maintenance in more direct ways, e.g. through metabolism, and which controls and coordinates the activities of other systems. Thus, they would deny that *E. coli* engages in cognition.

However, they accept that it engages in "intelligent behavior" (Moreno and Mossio 2015), and they can accept that it exhibits intentionality and has contentful states. Thus, for all they say, the intentionality of states of the nervous system is continuous with the intentionality of some states of other systems. Thus, intentionality may still be intimately linked to self-maintenance such that the default for intentional states is to motivate behaviour if they directly concern the organism's self-maintenance.

One may also worry that even if *E. coli* and similar simple organisms have *intentional* states, these states do not have *content*. Thus, the simple connection between content, self-maintenance and motivation would not hold. Hutto and Myin (2013, 2017), for instance, disagree with the form of autopoetic enactivism I am relying on. They defend a position which they call "Radical Enactive (or Embodied) Cognition" or "REC". They consider themselves in agreement with "autopoetic enactivism" and merely chide Thompson for using expressions like "meaning" and "significance", which they think is "misleading to analytic philosophers".

However, Hutto and Myin also hold that the cognitive activity of *E. coli*, and in fact all non-linguistic cognitive activity is not contentful. It does not involve representations or states with any 'content', including truth-conditional content. This, so Hutton and Myin (2013), is because any such account fails to resolve the "Hard Problem of Content". It cannot explain how contents fit into a world as understood by the (non-psychological) natural sciences. The only accounts which can resolve the

"Hard Problem of Content" rely on language and social practices. If Hutton and Myin's argument is successful, this would disprove the Ambitious Account. The Ambitious Account requires simple states, such as the states of *E. coli* to have truth-conditional content. If the simple states of *E. coli* did not have truth-conditional content, then the same

Hutto and Myin (2013) follow a 'divide-and-conquer' tactic, in which they criticise several different attempts to naturalise 'content'. They do not directly target E. Thompson or "autopoetic enactivism", as they take it to be a form of REC. However, they do attack "teleosemantics", and their arguments also apply to my 'REC-less' reading of Thompson. Firstly, they suggest that the notion of 'information-as-covariance' is simply too different from the notion of content to supply an account of truth-conditions, even when combined with a notion of proper function. Secondly, they argue that function-based accounts of content cannot account for "intensionality (with an *s*)".

By 'information-as-covariance', they refer to the notion also known as "Shannon information". This can be glossed, roughly, as the lawlike covariance between two properties such that an agent could use one to draw inferences about the other. In this sense, the rings of a tree carry 'information' about its age. Such information-of-covariance cannot be incorrect. If, due to some strange climatic occurrences, say, the rings of a tree are misleading concerning its age, they simply fail to carry information about its age. They do not carry false or incorrect information. Thus, information-as-covariance does not give rise to truth-conditions. Hutton and Myin insist that this does not change by adding the 'ingredient' of proper function. For example, they point out that even if a component of a system is *designed to* covary with another part of the system, it does not follow that it has truth-conditions.

However, not every account of content based on notions of biological function depends on information-as-covariance. Millikan's sender-receiver teleosemantics, importantly, does not depend on a notion of information-as-covariance. Millikan does *not* claim that contentful representations are states whose function it is to carry information. Rather, her view depends merely on a notion of *mapping* or *isomorphism*. For every signal (or contentful representation) there must be a sender and receiver. The sender and the receiver have complementary proper functions. The sender has the functions to produce states ('signals') which correspond to aspects of reality according to some definable isomorphism; the receiver has a proper function which it typically can only realise if the

signal corresponds to reality according to that isomorphism. This definition makes no reference to "information-as-covariance", but only to isomorphism.

Millikan's account also works for my 'REC-less' reading of autopoetic enactivism. Some subsystems within *E. coli* have the organisational function to produce internal states which are isomorphic to the environmental sugar-gradient. Other systems, the receiver systems, have the function of responding to this signal by moving. This constitutes intelligent action. Prototypical signals under this account will both have truth-conditions that involve the self-maintenance of the organism, and they will produce behaviour and thus potentially action.

Secondly, Hutto and Myin (2013) criticise accounts of content based in biological function for failing to capture how content exhibits "intensionality (with an *s*)". By "intensionality (with an *s*)", Hutto and Myin refer to the apparent phenomenon that contents consist not only of truth-conditions but also 'modes of presentations'. Thus, the beliefs that George Eliot is dead and that Mary Ann Evans is dead may have two different contents, yet the same truth-conditions. Hutto and Myin argue that 'intensionality (with an *s*)' can only be explained by reference to language and social practice. This may be true. However, Hutto and Myin do not provide an argument that all contentful states must be 'intensional (with an *s*)'. One might think that 'intensionality (with an *s*)' is necessary to resolve Frege puzzles, as described above. However, there might be intelligent systems with contentful states whose cognition is too simple to give rise to Frege puzzles. *E. coli* may be an example of such a creature. Insisting that the states of such a creature do not have truth-conditional content is merely question-begging. Hutto and Myin appear to think that 'intensionality (with an *s*)' is the mark of content: the way to distinguish contentful states from other states. However, there are plausible weaker criteria that may serve as a 'mark of content'. This is the case, for example, with *intentional inexistence*. Contentful states may represent objects that do not actually exist or states of affairs that do not obtain.

2.3 What about motivationally inert judgements about my own traits' organisational functions?

Doesn't the Ambitious Account prove too much? It must be the *truth-conditional content* which explains the motivational force of ethical judgements. If their motivational force was not explained

solely by this content, this would open the door to forms of mechanistic belief/desire-psychology. Thus, it would introduce the need to explain why agents have a separate desire for virtue. As long as I can have a belief with the same content as the motivationally efficacious judgement and not be motivated, it seems that there are really two mental entities which are, in principle, separable, even if they are somehow deeply connected. One is a conative state and the other cognitive. Hence, for the Ambitious Account to succeed, it must seemingly be the case that *any* judgement about my own traits' organisational functions would motivate me.

This is clearly false. An agent familiar with the organisational account may make a judgement they would express as 'If eaten, these fruits would contribute to my digestive system's organisational functions' without being in any way motivated to eat them. Similarly, as we have seen above. an agent who accepts that character traits have organisational functions, but who is not convinced of the organisational account of virtue could make a judgement expressible as 'If I was more daring, my character trait of courage would actualise its organisational function; as things are, it is malfunctioning.' Such an agent may well not be motivated in any way to be or become more daring. It is no more necessary that he should be motivated in line with that judgement than an agent who judges 'If I was more jealous, my character trait of sexual suspicion would realise its *selected-effect* function' must be motivated to be(come) jealous. This includes the conclusions of the external arguments for universal virtues discussed in Chapter Six. Even though they concern organisational functions, they cannot motivate without corresponding arguments internal to ethical inquiry.

The Default Account explained this by suggesting that the judgement in question presents the organisational function of character traits under a different mode of presentation as the (assumed) desire for virtue. According to the Ambitious Account, the motivational force of ethical judgements is explained by their *truth-conditional* content. Representations that differ only in mode of presentation have the same truth-conditional content. Hence, more must be said in order to show that the Ambitious Account can explain the same phenomenon. In the following, I will argue that there are certain ways in which an agent may make a judgement that do not exhibit *imaginative engagement*. Because they don't involve imaginative engagement, they do not generally motivate. The first of these ways consist in meta-linguistic judgements; the second consist in third-personal judgements about oneself. This solution is not incompatible with the solution adopted for the Default Account.

Meta-linguistic judgements and third-personal judgements can be understood as judgements under particular modes of presentation.

2.3.1 Meta-linguistic judgements

As for the first way, agents may judge that a certain linguistic statement is true without *imaginatively engaging* with its content, and thus without being motivated even if the statement concerns self-maintenance. However, if they are sufficiently competent with the expressions and constructions in the statement, they plausibly count as making a judgement whose content is that of the statement. Hence, they may make a judgement with a truth-conditional content concerning self-maintenance without being motivated.

To see this, consider first an agent judging that the following statement is currently true evaluated from their standpoint (in English): 'a dangerous ball is moving towards me'. Their attention is fully directed at the linguistic statement: at the words 'a', 'dangerous', 'ball', etc. Intuitively, it is much easier to imagine such an agent not being motivated to dodge the ball as it is with the agent described in 2.1 who sees the ball and forms the perceptual judgement that a dangerous ball is moving towards them. I suggest that we can say that the latter judgement involves *imaginative engagement* whereas the former doesn't. *Imagination*, here, should be understood in a sense where it is involved in ordinary perception. Judgements not involving imaginative engagement do not necessarily motivate action, even if they concern self-maintenance. However, imaginative engagement is not some additional state added onto our judgements. Rather, it is an intrinsic part of most of our ways of making judgements. The ways of making judgements without imaginative engagement, such meta-linguistic judgements are secondary and derivative.

Nonetheless, under the right conditions, the meta-linguistic judgements is true in all and only those possible worlds where a non-meta-linguistic judgement is true that could be expressed using the statement judged to be true. This is the case if the reference to the language of the is 'rigidified'. Even if the truth of the judgement is evaluated relative to a possible world in which the English language is different (say, 'ball' means *cream pie*), the relevant language is English as it is in the actual world. Plausibly, agents can only make such judgements if they are competent in the language in question. It is as if the English language is 'built into' the judgement. Thus, the meta-linguistic

judgement concerning the dangerous ball has the same truth-conditional content as the ordinary, non-metalinguistic judgement concerning it. Nonetheless, the agent is not motivated by it. Given their competence, they could, of course, transition to a judgement which involves imaginative engagement and thus motivates. However, doing so is not necessary or always automatic

This can be understood in a way that is compatible with autopoetic enactivism. According to autopoetic enactivism, intentionality is founded on organisms' interactions with their environment through which they self-maintain. With meta-linguistic judgements, however, the relevant intentional objects at which the judgement is immediately directed are linguistic objects, i.e. words. They are a feature of the social environment. However, these linguistic objects do not immediately call for any action. There is no way to interact with these *words* that would benefit self-maintenance when a ball is hurling at you. Hence, the judgements can be motivationally inert.

This contrast between purely meta-linguistic (or more generally, 'meta-representational') judgements and judgements involving imaginative engagement is not unique to judgements concerning self-maintenance. For example, if I judge that 5 + 2 = 7, I may do so in (at least) two ways. I may simply be responding automatically and habitually to the symbols used to express this judgement, e.g. the words *five, plus*, *two*, *equals* and *seven* or the characters <5>, <+>, <2>, <=> and <7>. Alternatively, I may, however vaguely, imagine five units (pebbles, fingers, dots, …) being added to two units. This difference is phenomenologically clear. It is also reflected in measurable evidence (cf. Gordon 2004, Sixtus et al. 2023). Thus, the appeal to the lack of imaginative engagement in order to explain some judgement's motivational inertness is not *ad hoc*.

Accounts of virtue and other ethical notions offered in views like bio-constructivism are complex enough that it is very difficult to make judgements concerning them except through meta-linguistic affirmation. Furthermore, the concepts used in them might be unfamiliar enough that it is very difficult to connect them with any imaginative engagement. Thus, it may plausibly be the case that the judgements we express using the terms of the account are typically only directed at the linguistic items expressing them. This includes the external arguments for universal virtues. Hence, just like in the case discussed above, they can be motivationally inert. This explains why it does not seem to us that such judgements could motivate us to action without the contribution of conative attitudes, even though they have the same truth-conditional content as virtue-judgements.

2.3.2 Third-personal judgements about oneself

Besides being cases of meta-linguistic affirmation, seeming ethical judgements may also be motivationally inert because the agent represents themselves in them as if they were someone else, a third person. In such cases, the agent does not imaginatively engage with the fact that he himself is the organism whose self-maintenance is under consideration.

Let us consider another example involving a dangerous ball. Anu sees a dangerous ball moving towards her in a mirror. Yet, she does not recognise that she is the person in the mirror. (Perhaps it is a funhouse mirror.) She forms a judgement she would express as: 'a dangerous ball is about to hit that person'. Suppose the expression 'that person' is purely referential. Insofar as it contributes to the truth-conditional content of the judgement, its meaning only consists of the referent, i.e. Anu. Thus, the judgement has the same truth-conditional content as the judgement Anu would express as 'I am about to be hit by a dangerous ball.' The contents of the two judgements are true in all and only the same possible worlds. Nonetheless, we would not expect the agent to be motivated to dodge the ball. This is compatible with autopoetic enactivism. Even though the judgement's truth-conditions are identical with those of a first-personal or *de se* judgement, the judgement does not intentionally disclose the environment of the agent (i.e. the ball) *as* her environment.

In Anu's case, her failure to connect the content of her judgement properly to herself is due to (non-culpable) ignorance. However, similar failures may also be due to insincerity or denial. Consider an agent who knows that if you are a smoker, you are likely to develop serious diseases. They apply this general statement to themselves and conclude that if they smoke, they will likely develop serious diseases. They are a smoker. Yet, they experience little to no motivational pull to quit smoking. In such a case, it is plausible to say that they do not properly connect their judgement to themselves. Like Anu, they judge about themselves as if they are judging about a third person. Another plausible example of this phenomenon are judgements about our own mortality. We all know that we will die. Yet, we struggle to picture a world in which we are dead. If I try to do so, I picture a world in which the person who bears my name, inhabits my body and has relations with my friends is dead. Yet, in this world there is still a point of view for 'me' to observe it from. I picture my death like the death of a third person. I do not imaginatively engage with the fact that it is *me* who will die.

The same may be the case for some ethical judgements. I may judge that it would be courageous for me to help my friend. However, I make this judgement in the same way as I would make a judgement about a third person. Thus, I am not motivated at all to help my friend. Plausibly, the concepts used in naturalistic analyses are especially likely to encourage such a third-personal perspective. The concepts are derived from natural science. Natural science is based on the exclusion of first-personal perspectives in favour of third-personal ones.

3. The normative challenge

In the preceding section, I have argued that ethical judgements and virtue-inquiry as conceived by bio-constructivism can motivate action. In this section, I focus on related objections that it is not *rational, reasonable* or *appropriate* for agents to be motivated by such judgements or that they *ought* not be motivated by such judgements in some other sense. There is a stronger and weaker version of this challenge. According to the strong challenge, agents ought to not be motivated by ethical judgements. It is irrational, unreasonable and/or inappropriate for them to be motivated in that way. According to the weaker challenge, it merely fails to be the case that agents generally ought to be motivated by ethical judgements as conceived by bio-constructivism. Being motivated by them is normatively neutral.

Both the weak and the strong challenge provide problems for bio-constructivism. According to bio-constructivism, virtue-inquiry successfully regulating character traits and motivating actions makes an agent *virtuous*; their behaviour comes closer to how they *ought* to act. If the strong challenge succeeds, this could not be so. Virtue-inquiry regulating action would make an agent *less* reasonable. Even if only the weaker challenge succeeds, an agent could not be criticised for being *insufficiently* motivated by their ethical judgements. We have already seen that according to the Default Account, an agent who is not motivated by their ethical judgements at all cannot be ethically evaluated. According to the Ambitious Account, such an agent is impossible. However, we can still criticise agents for failing to be sufficiently motivated by their virtue-inquiry, and acting viciously despite their better judgement. However, if being motivated by ethical judgements is normatively neutral, then presumably such motivation being more or less powerful is itself normatively neutral. Hence, an agent could not be criticised for being 'insufficiently' motivated by their ethical judgements. In

other words, regulation 'succeeding' or 'failing' would be completely neutral. Even beyond instances of ethical self-control, this makes the contention implausible that regulated-for traits are virtuous and regulated-against ones are vicious.

In the following, I will review a few versions of such 'normative' challenges to ethical self-control. Then, I will present a response to them.

3.1 Introducing the normative challenge

Firstly, Lott (2014) presents the "authority-of-nature challenge". A similar challenge has been independently described by Frey (2018) and De Anna (2018) as the "'irrelevance of nature' problem". However, I will focus on Lott's presentation. Lott notes that practical reason seems to make it possible for us to "'step back"' from desires and impulses and ask if they really provide a reason to act. As Lott recognises, this also applies to ethical considerations as described by neo-Aristotelian naturalism. According to neo-Aristotelian naturalism, virtues are traits of reason which human beings need in order to attain natural goods in a way that is characteristic of the human life-form. Using our ability to 'step back' from considerations, we can ask if we have a reason to live according to the human life-form. Neo-Aristotelian naturalism only describes what counts as good within the life-form. It does not show that this life-form as such is good. Thus it does not show that there is a reason to live according to it.

One might argue that for reason to even ask such a question is naturally defective. In order to play its characteristic role in human life, reason should not consider whether we have reason to live a characteristic human life. However, as a reason, this response is question-begging. It only provides a reason if we have a reason to live a characteristically human life. Furthermore, asking such questions follows from the internal characteristic of reason. Reason is the ability to treat considerations as reasons. This includes the ability to ask why we should treat a given consideration as a reason. A consistent understanding of a rational life-form must include this ability, rather than dismissing it as defective in some cases.

This challenge seemingly applies to bio-constructivism as much as to neo-Aristotelianism. Apparently, I am able to 'step back' from considerations of organisational function and ask if they

should really determine my actions. If I accept the organisational account of virtue, I can ask such a question regarding ethical judgements. After all, I believe that judgements of virtue-inquiry concern organisational function. If I cannot give some independent reason for letting myself be motivated by such considerations, then me being motivated by them is at best normatively neutral, as in the weak challenge and at worst unreasonable or irrational.

The authority-of-nature challenges resembles other challenges to neo-Aristotelian naturalism and to related views. David Enoch (2011) argues that normative facts, including moral facts, are "just too different" from natural facts to be identical with them. This includes facts about natural excellences as conceived by the neo-Aristotelian. Enoch presents this view at the end of his "indispensability argument". According to this argument, we are justified in believing in (robustly realist) normative facts because such facts are indispensable to our activity of practical reasoning. In this practice, argues Enoch, we are aiming at the truth about what we ought to do. Thus, for Enoch, facts about natural excellences are "just too different" from the facts we are aiming to adhere to in practical reasoning. Enoch himself confesses to being unable to clarify completely what the "just too different" intuition consists in. However, one plausible interpretation is this. Even if I accept that acting in a certain way would be naturally excellent, we can ask ourselves in practical reasoning if this is the way in which I ought to act. In other words, I can 'step back' from natural excellence. Thus, Enoch's worry turns out to be closely related to Lott's.

Similarly, Korsgaard (2019) argues that virtues as conceived by Foot and modern virtue ethicists (though not as conceived by Aristotle) are not constitutive of human agency. One can be fully a human agent (if in some sense a 'bad' one) while lacking these virtues. The appeal to 'constitutive standards' of agency is Korsgaard's (1996) own response to a similar challenge: How moral obligation can survive reflection about its sources. In other words, how can moral obligation be such that an agent cannot 'step back' from it?

3.2 The first, internal and holistic response

I will propose two responses to this objection. The first is 'internal' and based in the logic of ethical inquiry described in Chapter Five. The second is more speculative and based on the Ambitious Account presented in this chapter. It is meant to supplement, not replace the first response.

Firstly, as I have argued in Chapter Three much of the intuitive force behind arguments such as Enoch's Just-Too-Different Argument and Korsgaard's "normative question" is due to a worry about 'hypostatisation'. This is a worry about (meta-)ethical accounts according to which rational or ethically exemplary agents treat certain considerations as reasons for action even though they do not have an ethical reason for doing so. Thus, a certain way of living is removed from the scope of ethical critique. It becomes impossible to 'step back' from this way of life and subject it to critical ethical inquiry. However, bio-constructivism is not hypostatising in this sense. Ethical inquiry proceeds autonomously, based on its own logic and its own presuppositions. Every ethical judgment is supported by other ethical judgements in a holistic web. Furthermore, every ethical judgement can be questioned in ethical inquiry.

This includes also the judgement that one ought to be motivated by ethical judgements, i.e. that the 'desire for virtue' is virtuous. It is very likely that most or all agents whose ethical inquiry is functioning well would converge on the judgement that the desire for virtue is itself virtuous.

Firstly, just as I argued in Section 1 that there are evolutionary reasons to presume that most human agents find themselves with a desire for virtue, there are similar evolutionary reasons to presume that human agents generally find themselves with the prior judgement that the desire for virtue is virtuous. The desire for virtue enables self-regulation through virtue-inquiry. Such self-regulation benefits individual self-maintenance and thus fitness, as well as the survival of the individual's social group. However, this desire is also substantially benefited by the ethical judgement that the desire for virtue is virtuous. This judgement would regulate in favour of other subsystems which enhance the capacity of ethical inquiry to regulate character traits. In contrast, the alternative judgements that the desire for virtue is vicious or that is indifferent would undermine ethical self-regulation. Thus, the judgement that the desire for virtue is virtuous is likely to be selected for and/or inculcated into children in their upbringing. Hence, through the constraint of diachronic consistency, it is likely to be endorsed in maximally functional virtue-inquiry. Furthermore, the desire for virtue is likely to increase how much virtuous acts compared to vicious acts an agent performs. It shares this feature with most virtues but not with most vices or indifferent traits. Hence, the constraint of synchronic consistency will count in favour of judging the desire for virtue virtuous.

The arguments within virtue-inquiry for the conclusion that the desire for virtue is virtuous are independent of the organisational account of virtue. However, they are compatible with it. They are not undermined by accepting that virtuousness is identical with the organisational function of character traits.

Consequently, the ethically exemplary agent according to bio-constructivism is not motivated by their ethical judgements without being able to 'step back' from such a motivation. Rather, they can critically inquire concerning whether such motivation is virtuous. They will (almost certainly) find that it is.

3.3 The second, external response

However, one may find the internal response to be unsatisfying. After all, it may not be tautological that maximally functional virtue-inquiry as conceived by bio-constructivism endorses the disposition to act according to precepts of maximally functional virtue-inquiry as conceived by bio-constructivism, but it is uncomfortably circular. A doubter who questions whether virtue-inquiry as conceived by bio-constructivism has any normative claim on her actions is unlikely to be convinced by it. They demand a wholesale, external validation of virtue-inquiry and its normative claim. In the section, I argue that the Ambitious Account can plausibly provide such a response. Given that the Ambitious Account is itself speculative, this response will itself be speculative, and I will only outline it briefly. If autopoetic enactivism and the Ambitious Account are correct, then being motivated by approximately maximally functional virtue-inquiry is necessary for practical reason to have intentionality. It is necessary for practical reason to be about anything at all.

According to autopoetic enactivism, other things being equal, an intentional representation with a truth-conditional content concerning organisational functions will motivate its bearer to behave in certain ways. Hence, if an agent apparently makes the judgements that are distinctive of virtue-inquiry according to bio-constructivism, but they are not motivated by them, then this means that the apparent judgements are not intentional representations. Given that virtue-inquiry takes a central role in ethical inquiry and practical reason, if virtue-inquiry does not consist of intentional representations, then other sophisticated and conceptual components of practical reason will not feature intentional representations either. Some emotional representations which are part of practical reason, i.e. the human capacity to treat considerations as reasons, may still have intentional content

due to their connection to action. However, more abstract and sophisticated components of practical reason are only connected to action through virtue-inquiry. The capacity through which a doubter may ask whether virtue-inquiry as conceived by bio-constructivism has any normative claim on her is one such abstract and sophisticated component. Hence, if the agent is not motivated to act according to virtue-inquiry, then the apparent thought through which the doubter asks themselves whether to act in the way virtue-inquiry as conceived by bio-constructivism has no intentional content. In even asking themselves the question, the doubter assumes that it has some kind of content. Otherwise, there would be no way of answering it. Hence, in even asking the question, the doubter is committed to a positive answer.

This account resembles other so-called constitutivist accounts of normativity. These are accounts which attempt to explain the fact that agents ought to be P by arguing that P is necessary for being an agent (e.g. Korsgaard 2009). Enoch (2006, 2011) has argued that such accounts fail to respond to the doubter's worry about whether to follow the precepts of normativity. An agent may simply say: 'If I don't follow the precepts of normativity, I'm not an agent. So what! I'll be a shmagent instead.' A shmagent is like an agent except they don't follow normative precepts.

However, unlike some other constitutivist accounts, the Ambitious Account connects ethical motivation to the overtly non-practical notion of intentionality. Through it, it is connected to truth. Plausibly, representation is true iff it accurately represents that which it is about. Hence, an agent who presumed that they might be a shmagent would also have to presume that the states of their practical reason exhibit 'shmintentionality' and that they cannot be true but only 'shmue'. Plausibly there are no such coherent alternatives to the notions of intentionality and truth.

Of course, this second response would not convince the inveterate amoralists. It could not even serve to convict her of (non-practical) irrationality. For one, the Ambitious Account is contentious and depends on empirical knowledge to make it plausible. Hence, amoralists could faultlessly reject that account, and thus a necessary premise of that account. However, for the agent who is convinced by Ambitious Account, the secondary response provides a way to reassure themselves if they doubt whether to act in line with virtue-inquiry as conceived by bio-constructivism and a way to integrate ethical motivation further into their web of beliefs and commitments.

4. Conclusion

In this chapter, I have provided two accounts of how ethical inquiry as conceived by bio-constructivism may motivate action, as is required for it to regulate character traits. The Default Account simply assumes that agents have a desire for virtue which is connected to ethical inquiry. A more ambitious Ambitious Account instead explains such a desire through an account of intentional representation.

EIGHT: Conclusion

In this thesis I have introduced a novel form of function-based naturalism. This account identifies virtues as character traits actualising their regulated functions as described by the organisational account of function. Character traits are regulated by virtue-inquiry, i.e. that component of ethical inquiry concerned with which considerations to treat as reasons for action. I have not attempted to provide conclusive arguments for this argument. However, I have attempted to show that it is viable, interesting, and avoids serious objections that can be levelled against other forms of evolutionary and neo-Aristotelian function-based naturalisms.

This account avoids both the challenge from science and the hypostatisation challenge. It avoids the hypostatisation challenge because it does not place any way of being human beyond ethical criticism or assumes that it is virtuous without providing ethical reasons. This is because the regulated functions of character traits are largely determined by virtue-inquiry as a general regulatory system. All kinds of incompatible character traits could contribute to self-maintenance on different levels of self-maintenance (the level of mental organisation, the individual, the narrow lineage, the extended lineage, etc.). Virtue-inquiry determines which actualises a regulated function. Virtue-inquiry does not itself have a regulated function; however it does have a bare function. It contributes to self-maintenance in a consistent and systematic way. It is only genuinely regulating character traits if it actualises that function. Thus, in order to discover the regulated functions of character traits, we must let virtue-inquiry 'play out'. It follows its own logic. Within virtue-inquiry, any way of being human can be criticised and the criticism evaluated; any judgement must be supported by ethical reasons.

At the same time, the account can avoid the challenge from science. Virtue-inquiry is not a superstitious or pseudo-scientific practice in which the conclusions reached are unrelated to the truth in the subject matter. Rather, character traits have regulated functions in exactly the same way as other human traits and traits of other living beings. Judgements in virtue-inquiry are true (or false) because of how they (fail to) correspond to the reality they purport to describe in just the same way as other judgements in biology. Furthermore, given that it is not accidental for virtue-inquiry to actualise its function, there is a connection between the reality described by ethical judgements and the judgements themselves.

Bio-constructivism's capacity to realise the naturalistic strategy in responding to the challenge from science while avoiding hypostatisation constitutes the core of its appeal. This appeal is supported by its capacity to explain particular features of the distinctive logic of ethical inquiry (Chapter Five), to let us derive plausible universal virtues while seeing continuities between human character traits and traits of other living beings in epistemically fruitful ways (Chapter Six) and to provide a speculative but potentially very elegant account of ethical motivation (Chapter Seven)

References

Adam, T. C., & Epel, E. S. (2007). Stress, eating and the reward system. *Physiology & Behavior, 91*(4), 449-458.

Adorno, T. W. (1966). *Negative Dialektik*. Suhrkamp.

Alcoff, L. M. (2006). *Visible identities : race, gender, and the self*. Oxford University Press.

Alvard, M., & Nolin, D. (2024). Cooperation. In: J. Koster, B. Scelza, & M. K. Shenk (Eds.), *Human Behavioral Ecology* (pp. 104–129), Cambridge: Cambridge University Press.

Anderson, E. S. (1991). John Stuart Mill and experiments in living. *Ethics, 102*(1), 4-26.

Andreou, C. (2006). Getting on in a Varied World. *Social Theory and Practice, 32*(1), 61-73.

Andrews, D., Lala, K. N., Uller, T., Feiner, N., Feldman, M., & Gilbert, S. F. (2024). *Evolution Evolving : The Developmental Origins of Adaptation and Biodiversity.* (1st ed.). Princeton University Press. https://doi.org/10.1515/9780691262406

Annas, J. (2008). Virtue ethics and the charge of egoism. In: P. Bloomfield (Ed.). *Morality and self-interest* (pp. 205-223). Oxford: Oxford University Press

Anscombe, G. E. M. (1978). Rules, Rights and Promises. *Midwest Studies in Philosophy, 3*(1), 318-323.

Aquinas, T. , & McDermott, T. S. (1991 [1274]). *Summa theologiae : a concise translation*. Methuen.

Armitage, D., & Gaffield, J. (2016). The Haitian Declaration of Independence in an Atlantic Context. In J. Gaffield (Ed.), *The Haitian Declaration of Independence: Creation, Context, and Legacy* (pp. 1–22). University of Virginia Press.

Arnellos, A., & Moreno, A. (2022). Cognitive functions are not reducible to biological ones: the case of minimal visual perception. *Biology & Philosophy, 37*(4), 35.

Arnaud, S., Sullivan, J., MacKinnon, A., & Bodell, L. P. (2024). Teasing apart the roles of interoception, emotion, and self-control in anorexia nervosa. *Review of Philosophy and Psychology, 15*(3), 723-747.

Arnhart, L. (1998). *Darwinian Natural Right : The Biological Ethics of Human Nature*. State University of New York Press.

Ayer, A. J. (2012 [1952]). *Language, Truth and Logic*. Dover Publications.

Battivelli, D., Fan, Z., Hu, H., & Gross, C. T. (2024). How can ethology inform the neuroscience of fear, aggression and dominance?. *Nature Reviews Neuroscience, 25*(12), 809-819.

Bell, M. (2009). Anger, Virtue, and Oppression. In Tessman, L. (ed.) *Feminist Ethics and Social and Political Philosophy: Theorizing the Non-Ideal* (pp. 165–183). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6841-6_10

Bickhard, M. H. (2000). Autonomy, function, and representation. *Communication and Cognition-Artificial Intelligence, 17*(3-4), 111-131.

Bickhard, M. H. (2004). Part II: applications of process-based theories: process and emergence: normative function and representation. *Axiomathes, 14*(1-3), 121-155.

Bickhard, M. H. (2009). Interactivism: A manifesto. *New Ideas in Psychology, 27*(1), 85-95.

Bickhard, M. H. (2024). *The Whole Person : Toward a Naturalism of Minds and Persons*. Academic Press.

Bigelow, J., & Pargetter, R. (1987). Functions. *The Journal of Philosophy*, *84*(4), 181–196. https://doi.org/10.2307/2027157

Bookchin, M. (1982). *The ecology of freedom : the emergence and dissolution of hierarchy*. Cheshire Books.

Boorse, C. (1976). Wright on functions. *The Philosophical Review, 85*(1), 70-86.

Boorse, C. (2014). A second rebuttal on health. *Journal of medicine and philosophy, 39*(6), 683-724.

Boyd, R. (1988). How to Be a Moral Realist. In: Sayre-McCord, G. (ed.). *Essays on moral realism* (pp. 181-228). Cornell University Press.

Boyd, R. (1999). Homeostasis, Species, and Higher Taxa. In: Wilson, R. A. (ed.). *Species : new interdisciplinary essays* (pp. 141-85). MIT Press.

Broad, C. D. (1944). Huxley, J. S. - Evolutionary Ethics [Review]. *Mind, 53*, 344.

Brink, D. O. (1984). Moral realism and the sceptical arguments from disagreement and queerness. *Australasian Journal of Philosophy, 62*(2), 111-125.

Bryson, S. T. (2013). The Art of Power: Poison and Obeah Accusations and the Struggle for Dominance and Survival in Jamaica's Slave Society. *Caribbean Studies, 41*(2), 61-90.

Callicott, J. B. (1982). Hume's Is/ought Dichotomy and the Relation of Ecology to Leopold's Land Ethic. *Environmental Ethics, 4(*2), 163-174.

Campbell, R. (1996). Can biology make ethics objective?. *Biology and Philosophy, 11*(1), 21-31.

Carus, P. (1899). Immortality as a Philosophic Principle. *The Monist*, *9*(4), 572–616.

Chappell, T. (2011). Glory as an ethical idea. *Philosophical Investigations, 34*(2), 105-134.

Chappell, S.G. (2015). Lists of Virtues. *Etica & Politica / Ethics & Politics*, *17,* 74-93

Chappell, S.G. (2023). Tempted Like Achilles: Reflections on Roles and Role-Recalcitrance in Ancient and Modern Ethics. In: Barber, A., Cordell, S., Barber, A., & Cordell, S. (eds.). *The Ethics of Social Roles* (pp. 25–48). Oxford University Press.

Chappell, S.G. (2024). Utrum sit una tantum vera enumeratio virtutum moralium (*Whether There Is a Single Correct List of the Virtues of Character*). *New Blackfriars, 105*(5), 471-477.

Cherry, M. V. (2021). *The case for rage : why anger is essential to anti-racist struggle*. Oxford University Press.

Chisholm, J. S. (1999). *Death, hope and sex : steps to an evolutionary ecology of mind and morality*. Cambridge University Press.

Christensen, W. D., & Bickhard, M. H. (2002). The process dynamics of normative function. *The Monist, 85*(1), 3-28.

Churchland, P. M. (1989). *A neurocomputational perspective : the nature of mind and the structure of science*. MIT Press.

Collier, J., & Stingl, M. (1993). Evolutionary naturalism and the objectivity of morality. *Biology and Philosophy, 8*(1), 47-60.

Collier, J. (2000). Autonomy and process closure as the basis for functionality. *Annals of the New York Academy of Sciences, 901*(1), 280-290.

Corrington, G. (1986). Anorexia, asceticism, and autonomy: Self-control as liberation and transcendence. *Journal of Feminist Studies in Religion, 2*(2), 51-61.

Darwin, C. (1910 [1859]). *Origin of the species by means of natural selection*. J. Murray.

Dawkins, R. (1993). Viruses of the Mind. In: Dahlbom, B. (ed.). *Dennett and His Critics : Demystifying Mind* (pp. 13-27). Blackwell.

De Jong, D. C. (2009). The role of attention in sexual arousal: Implications for treatment of sexual dysfunction. *Journal of Sex Research, 46*(2-3), 237-248.

De Roode, J. C., Pansini, R., Cheesman, S. J., Helinski, M. E., Huijben, S., Wargo, A. R., ... & Read, A. F. (2005). Virulence and competitive ability in genetically diverse malaria infections. *Proceedings of the National Academy of Sciences, 102*(21), 7624-7628.

de Waal, F., & Sherblom, S. A. (2018). Bottom-up morality: The basis of human morality in our primate nature. *Journal of Moral Education, 47*(2), 248-258.

Descartes, R. (1998 [1641]). *Discourse on method ; and, Meditations on first philosophy* (D. A. Cress, Tran.; Fourth edition.). Hackett Publishing Company.

Dosen, A. S., & Ostwald, M. J. (2016). Evidence for prospect-refuge theory: a meta-analysis of the findings of environmental preference research. *City, Territory and Architecture, 3*(1), 1-14.

Douglass, F. (2015 [1845]). *Narrative of the life of Frederick Douglass*. Vintage Digital.

Dove, M. (2024). *Hearsay is not excluded : a history of natural history*. Yale University Press.

Drummond, J. (2002). Aristotelianism and phenomenology. In: Drummond, J. & Embree, L. (eds.) *Phenomenological approaches to moral philosophy: A handbook* (pp. 15-45). Springer

Drummond, J. (2013). Phenomenology, Eudaimonia, and the Virtues. In: Gyllenhammer, P., & Hermberg, K. (eds.) *Phenomenology and Virtue Ethics* (pp. 97–112). Bloomsbury Academic. https://doi.org/10.5040/9781472548153.ch-006

Drummond, J. (2015). Neo-Aristotelian Ethics: Naturalistic or Phenomenological. *Phenomenology in a New Key: Between Analysis and History: Essays in Honor of Richard Cobb-Stevens*, 135-149.

Dussault, A. C. (2022). Two notions of ecological function. *Philosophy of Science, 89*(1), 171-179.

Dwyer, E. A. (2024). The Poison Pen: Slavery, Poison, and Fear in the Antebellum Press. In *Slavery and Emotions in the Atlantic World* (pp. 13-29). Routledge.

Efferson, C., Lalive, R., & Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science, 321*(5897), 1844-1849.

El-Hani, C. N., De Lima, F. R. G., & Nunes-Neto, N. (2024). From the organizational theory of ecological functions to a new notion of sustainability. In: Mossio, M. (ed) *Organization in Biology. History, Philosophy and Theory of the Life Sciences, vol 33* (pp. 285-328). Springer, Cham.

Enoch, D. (2006). Agency, Shmagency: Why normativity won't come from what is constitutive of action. *The Philosophical Review, 115*(2), 169-198.

Enoch, D. (2010). Shmagency Revisited. In: Brady, M. (Ed.) *New Waves in Metaethics* (pp. 208–233). Palgrave Macmillan UK.

Enoch, D. (2011). *Taking Morality Seriously : a Defense of Robust Realism*. Oxford University Press.

Foot, P. (1978). *Virtues and vices and other essays in moral philosophy*. University of California Press.

Foot, P. (2001) *Natural Goodness*. Clarendon Press.

Foster, K. R. (2004). Diminishing returns in social evolution: the not-so-tragic commons. *Journal of Evolutionary Biology, 17*(5), 1058-1072.

Frank, S. A. (1995). Mutual policing and repression of competition in the evolution of cooperative groups. *Nature, 377*(6549), 520-522.

Frank, S. A. (1996). Models of parasite virulence. *The Quarterly review of biology, 71*(1), 37-78.

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy, 68*(1), 5-20.

Freeman, T., Gesesew, H. A., Bambra, C., Giugliani, E. R. J., Popay, J., Sanders, D., ... & Baum, F. (2020). Why do some countries do better or worse in life expectancy relative to income? An analysis

of Brazil, Ethiopia, and the United States of America. *International Journal for Equity in Health, 19*, 1-19.

Freyenhagen, F. (2013). *Adorno's practical philosophy : living less wrongly*. Cambridge University Press.

Freyenhagen, F. (2023). The Linguistic Turn in the Early Frankfurt School: Horkheimer and Adorno. *Journal of the History of Philosophy, 61*(1), 127-148.

Fricker, M. (2020). Bernard Williams as a philosopher of ethical freedom. *Canadian Journal of Philosophy, 50*(8), 919-933.

Fu, F., Tarnita, C. E., Christakis, N. A., Wang, L., Rand, D. G., & Nowak, M. A. (2012). Evolution of in-group favoritism. *Scientific Reports, 2*(1), 1-6.

Garson, J. (2022). What are Functions Good For?. *Australasian Philosophical Review, 6*(4), 374–385. https://doi.org/10.1080/24740500.2024.2370624

Gelman, S.A. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford: Oxford University Press.

Gensler, H. J. (2013). *Ethics and the Golden Rule*. Routledge.

Gervais, M., & Wilson, D. S. (2005). The evolution and functions of laughter and humor: A synthetic approach. *The Quarterly Review of Biology, 80*(4), 395-430.

Gibbard, A. (2003). *Thinking how to live*. Harvard University Press.

Godfrey-Smith, P. (1996). *Complexity and the function of mind in nature*. Cambridge University Press.

Godfrey-Smith, P. (2016a). Individuality, subjectivity, and minimal cognition. *Biology & Philosophy, 31*(6), 775-796.

Godfrey-Smith, P. (2016b). Mind, matter, and metabolism. *The Journal of Philosophy, 113*(10), 481-506.

Godfrey-Smith, P. (2016c). Animal evolution and the origins of experience. In: Smith, D. L. (Ed.). *How biology shapes philosophy : new foundations for naturalism* (pp. 51-71). Cambridge University Press.

González de Prado Salas, J. (2018). Whose purposes? Biological teleology and intentionality. *Synthese, 195*(10), 4507-4524.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*(5695), 496-499.

Gowans, C. W. (2008). Virtue and Nature. *Social Philosophy and Policy, 25*(1), 28-55.

Hacker-Wright, J. (2021). *Philippa Foot's Metaethics*. Cambridge University Press.

Hadley, A. T. (1913). *Some influences in modern philosophic thought*. Yale University Press.

Haeckel, E. (1899). *Die Welträthsel: Gemeinverständliche Studien über monistische Philosophie*. Bonn: Strauß. https://www.deutschestextarchiv.de/book/show/haeckel_weltraethsel_1899

Harcourt, E. (2015). The place of psychoanalysis in the history of ethics. *Journal of Moral Philosophy, 12*(5), 598-618.

Hardin, G. (1968). The Tragedy of the Commons. *Science, 162* ,1243-1248.

Hare, R. M. (1952). *The Language of Morals*. Clarendon Press.

Heck, R. G. (2012). Solving Frege's puzzle. *The Journal of Philosophy, 109*(1/2), 132-174.

Henriques, G. (2003). The tree of knowledge system and the theoretical unification of psychology. *Review of General Psychology, 7*(2), 150-182.

Hiernaux, Q. (2023). *From Plant Behavior to Plant Intelligence?* Editions Quae.

Hollenstein, T. (2015). This time, it's real: Affective flexibility, time scales, feedback loops, and the regulation of emotion. *Emotion Review, 7*(4), 308-315.

Hooker, C. A. (2009). Interaction and bio-cognitive order. *Synthese, 166*, 513-546.

Hooker, C. (2024). On the Organizational Roots of Bio-cognition. In: Mossio, M. (ed.) *Organization in Biology. History, Philosophy and Theory of the Life Sciences, vol 33*. Springer, Cham. https://doi.org/10.1007/978-3-031-38968-9_5

Horkheimer, M. and Adorno, T. W. (1944). *Dialektik der Aufklärung : philosophische Fragmente*. S. Fischer.

Hume, D. (1740). *A Treatise of Human Nature*. https://davidhume.org/texts/t/

Hume. D. (1777) *Of the Populousness of Ancient Nations*. https://davidhume.org/texts/empl2/pa

Hurka, T. (2001) *Virtue, Vice and Value*. Oxford: Oxford University Press

Hursthouse, R. (1999) *On Virtue Ethics*. Oxford University Press

Hutto, D. D., & Myin, E. (2013). *Radicalizing Enactivism : Basic Minds Without Content*. MIT Press.

Hutto, D. D., & Myin, E. (2017). *Evolving Enactivism : Basic Minds Meet Content*. The MIT Press.

Huxley, J. (1943). *Evolutionary Ethics*. Oxford University Press, Humphrey Milford.

Huxley, T. H. (1894). *Evolution & Ethics and Other Essays*. Macmillan.

Jackson, F., & Pettit, P. (1995). Moral functionalism and moral motivation. *The Philosophical Quarterly, 45*(178), 20-40.

Jefferson, T. (2022 [1783]). *Notes on the State of Virginia* (R. P. Forbes, Ed.; Annotated Edition.). Yale University Press.

Kant, I. (2019 [1785]). *Groundwork for the Metaphysics of Morals* (C. Bennett, J. Saunders, & R. Stern, Trans.). Oxford University Press.

Keener, J., & Sneyd, J. (2009). *Mathematical Physiology II: Systems Physiology*. New York: Springer.

Keller, S. (2007). Virtue ethics is self-effacing. *Australasian Journal of Philosophy, 85*(2), 221-231.

Kelly, D., Machery, E, and Mallon, R. (2010). Race and Racial Cognition. In Doris, J.M. and Moral Psychology Research Group (eds.) *The Moral Psychology Handbook* (pp. 433–69). Oxford: Oxford University Press,

Kerr, B., Neuhauser, C., Bohannan, B. J., & Dean, A. M. (2006). Local migration promotes competitive restraint in a host–pathogen 'tragedy of the commons'. *Nature, 442*(7098), 75-78.

Kimble, M., Boxwala, M., Bean, W., Maletsky, K., Halper, J., Spollen, K., & Fleming, K. (2014). The impact of hypervigilance: Evidence for a forward feedback loop. *Journal of Anxiety Disorders, 28*(2), 241-245.

Kitcher, P. (1993). Function and Design. *Midwest Studies in Philosophy*, *18*(1), 379–397.

Korsgaard, C.M. (1985), Kant's Formula of the Universal Law. *Pacific Philosophical Quarterly, 66*. 24-47.

Korsgaard, C. M. (1996). *The Sources of Normativity*. Cambridge University Press.

Korsgaard, C. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.

Kropotkin, P. A. (2022 [1902]). *Mutual Aid : a Factor of Evolution*. Penguin Books.

Lakatos, I. (1968). Criticism and the Methodology of Scientific Research Programmes. *Proceedings of the Aristotelian Society*, *69*, 149–186.

Leopold, A. (2001 [1949]). *A Sand County Almanac : with Essays on Conservation*. Oxford University Press.

Levine, S., Kleiman-Weiner, M., Schulz, L., Tenenbaum, J., & Cushman, F. (2020). The logic of universalization guides moral judgment. *Proceedings of the National Academy of Sciences, 117*(42), 26158-26169.

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50(3), 249-258.

Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review, 88*(4), 513-543.

Lockhart, J. R., & Lott, M. (2024). Are life forms real? Aristotelian naturalism and biological science. *Synthese, 203*(3), 71.

Lott, M. (2020). Aristotelian Naturalism and the Autonomy of Ethics. In: Hähnel, M. (Ed.). *Aristotelian Naturalism* (pp. 283–294). Springer International Publishing.

Lux, K. (1990). *Adam Smith's Mistake : How a Moral Philosopher Invented Economics & Ended Morality*. Shambhala.

Machery, E., Faucher, L., & Kelly, D. R. (2010). On the alleged inadequacies of psychological explanations of racism. *The Monist, 93*(2), 228-254.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Penguin Books.

MacIntyre, A. C. (1981). *After virtue : a study in moral theory*. Duckworth.

MacIntyre, A. C. (1988). *Whose justice? Which rationality?*. Duckworth.

MacIntyre, A. C. (1999). *Dependent Rational Animals : Why Human Beings Need the Virtues*. Duckworth.

MacIntyre, A. (2002). Virtues in Foot and Geach. *The Philosophical Quarterly, 52*(209), 621–631.

Mallon, R. (2013). Was Race Thinking Invented in the Modern West?. *Studies in History and Philosophy of Science Part A, 44*(1), 77-88.

Malthus, T. R. (2008 [1798]). *An Essay on the Principle of Population* (G. Gilbert, Ed.). Oxford University Press.

Mann, S. F., & Pain, R. (2022a). Teleosemantics and the Hard Problem of Content. *Philosophical Psychology, 35*(1), 22-46.

Mann, S. F., & Pain, R. (2022b). Teleosemantics and the free energy principle. *Biology & Philosophy, 37*(4), 34.

Maturana, H., & Varela, F. J. (1980). *Autopoiesis and Cognition: the Realization of the Living*. D. Reidel Pub. Co.

McDowell, J. (1995). Might there be external reasons? In J. E. J. Altham & R. Harrison (Eds.), *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (pp. 68–85). Cambridge: Cambridge University Press.

McDowell, J. (1998). Two sorts of naturalism. *Mind, value, and reality*, 167-97.

McKibbin, W. F., Shackelford, T. K., Goetz, A. T., & Starratt, V. G. (2008). Why do men rape? An evolutionary psychological perspective. *Review of General Psychology, 12*(1), 86-97.

McLaughlin, P. (2002). On having a function and having a good. *Analyse & Kritik, 24*(1), 130-143.

McMullin, I. (2019). *Existential Flourishing : a Phenomenology of the Virtues*. Cambridge University Press.

McRae, K., & Gross, J. J. (2020). Emotion Regulation. *Emotion, 20*(1), 1.

Mercier, H., & Sperber, D. (2017). *The Enigma of Reason*. Harvard University Press.

Merleau-Ponty, M. (2012 [1945]). *Phenomenology of Perception* (D. A. Landes, Ed.). Routledge.

Midgley, M. (1979). *Beast and Man: The Roots of Human Nature*. Harvester.

Mill, J. S. (2014 [1859]). *On Liberty*. AUK Classics

Millikan, R. G. (1989). In Defense of Proper Functions. *Philosophy of Science, 56*(2), 288–302

Millikan, R. G. (2004). *Varieties of Meaning : The 2002 Jean Nicod Lectures*. MIT Press.

Millstein, R. L. (2024). *The Land Is Our Community : Aldo Leopold's Environmental Ethic for the New Millennium*. University of Chicago Press.

Millum, J. (2006). Natural Goodness and Natural Evil. *Ratio, 19*(2), 199-213.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 80(4), 252.

Mischel, W. (2009). From Personality and Assessment (1968) to Personality Science, 2009. *Journal of Research in Personality, 43*(2), 282-290.

Mix, L. (2020). Three Lives and Astrobiology. *Social and Conceptual Issues in Astrobiology, 57*

Mohamed, S. (2015). Of Monsters and Men: Perpetrator Trauma and Mass Atrocity. *Colum. L. Rev., 115*, 1157.

Moore, G. E. (1922). *Principia Ethica*. Cambridge University Press.

Moosavi, P. (2019). From Biological Functions to Natural Goodness. *Philosophers' Imprint* 19.

Moreno, Á. & Mossio, M. (2015). *Biological Autonomy : a Philosophical and Theoretical Enquiry*. Springer.

Mossio, M., & Bich, L. (2017). What makes biological organisation teleological?. *Synthese, 194*(4), 1089-1114.

Mossio, M., & Pontarotti, G. (2022). Conserving Functions across Generations: Heredity in Light of Biological Organization. *The British Journal for the Philosophy of Science, 73*(1).

Mossio, M., Saborido, C., & Moreno, A. (2009). An Organizational Account of Biological Functions. *The British Journal for the Philosophy of Science*, *60*(4), 813–841.

Nadkarni, M.V. (2014). Ethics in Hinduism. In *Ethics for Our Times*. Delhi: Oxford University Press.

Nagel, T. (1969). Sexual Perversion. *The Journal of Philosophy, 66*(1), 5-17.

Nagel, T. (1970). *The Possibility of Altruism*. Princeton University Press.

Neander, K. (2017). Functional analysis and the species design. *Synthese, 194*(4), 1147-1168.

Nietzsche, F. W. (2008 [1889]). *The case of Wagner, Nietzsche contra Wagner, and selected aphorisms* (A. M. (Anthony M. Ludovici, Tran.; 3rd edition.). Dodo Press.

Nietzsche, F. W., & Del Caro, A. (2012 [1883]). *Thus Spoke Zarathustra* (R. B. Pippin, Ed.). Cambridge University Press.

Nietzsche, F. W. (2016 [1878]). *Human, All-Too-Human: A Book for Free Spirits, Part 1 Complete Works, Volume Six* (O. Levy, Ed.; H. Zimmern, Tran.). Project Gutenberg.

Norton, B. G. (2011). What Leopold learned from Darwin and Hadley: comment on Callicott et al. *Environmental Values, 20*(1), 7-16.

Nozick, R. (1989). *The Examined Life*. Simon & Schuster

Nunes-Neto, N., Moreno, A., & El-Hani, C. N. (2014). Function in ecology: an organizational approach. *Biology & Philosophy, 29*, 123-141.

Nunes-Neto, N., do Carmo, R. S., & El-Hani, C. N. (2016a). Biodiversity and ecosystem functioning: An analysis of the functional discourse in contemporary ecology. *Filosofia e História da Biologia, 11*(2), 289-321.

Nunes-Neto, N., Saborido, C., El-Hani, C. N., Viana, B., & Moreno, A. (2016b). Function and normativity in social-ecological systems. *Filosofia e História da Biologia, 11*(2), 259-287.

Nussbaum, M. C. (1988). Non-Relative Virtues: An Aristotelian Approach. *Midwest Studies in Philosophy*, *13*(1), 32–53.

Odenbaugh, J. (2017). Nothing in ethics makes sense except in the light of evolution? Natural goodness, normativity, and naturalism. *Synthese, 194*, 1031-1055.

Osborne, C. R. (1974). *The Tiwi language : grammar, myths and dictionary of the Tiwi language spoken on Melville and Bathurst Islands, northern Australia*. Australian Institute of Aboriginal Studies.

Ostrom, E. (1990). *Governing the commons : the evolution of institutions for collective action*. Cambridge University Press.

Parfit, D. (1984). *Reasons and persons*. Clarendon Press.

Parfit, D., & Scheffler, S. (2011). *On what matters. Volume one* (1st ed.). Oxford University Press.

Parker, S. (1982, Winter). Ragnar Redbeard and Right of Might. *The Storm*. http://www.sidparker.com/essays/ragnar-redbeard-and-the-right-of-might/

Pettigrove, G. (2011). Is virtue ethics self-effacing?. *The Journal of Ethics, 15*, 191-207.

Pettigrove, G. (2012). Meekness and 'moral' anger. *Ethics, 122*(2), 341-370.

Pölzler, T., & Wright, J. C. (2020). Anti-Realist Pluralism: A New Approach to Folk Metaethics. *Review of philosophy and psychology*, 11(1), 53–82.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. Hutchinson.

Powell, R. (2023). Social norms and superorganisms. *Biology & Philosophy*, *38*(3), 21.

Prior, A. N. (1960). The Autonomy of Ethics. *Australasian Journal of Philosophy, 38*(3), 199-206.

Quine, W. V. (1951). *Two Dogmas of Empiricism*. Longmans, Green.

Rama, T. (2023). Evolutionary Causation and Teleosemantics. In: Viejo, J.M., Sanjuán, M. (eds) *Life and Mind. Interdisciplinary Evolution Research, vol 8*. Springer, Cham. https://doi.org/10.1007/978-3-031-30304-3_14

Rama, T. (2024). Is a Cognitive Revolution in Theoretical Biology Underway?. *Foundations of Science*, 1-22.

Rankin, D. J., Bargum, K., & Kokko, H. (2007). The tragedy of the commons in evolutionary biology. *Trends in Ecology & Evolution, 22*(12), 643-651.

Rankin, D. J., & Kokko, H. (2006). Sex, death and tragedy. *Trends in Ecology & Evolution, 21*(5), 225-226.

Redbeard, R. (1910 [1896]). *Might is right; or, The survival of the fittest.* W.J. Robbins.

Rifkin, M. J., & Garson, J. (2023). Sex by design: a new account of the animal sexes. *Biology & Philosophy, 38*(2), 13.

Robbins, P. (2017). Modularity of Mind. In Edward N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), URL = <https://plato.stanford.edu/archives/win2017/entries/modularity-mind/>.

Rodrigue, L. (2014). The Virtue of Gentleness in Aristotle‘s Ethics. *Dionysius, 32*, 33-53.

Ross, J. (2019). From Teleosemantics to Normative Ethics. In: M. Timmons (ed.), *Oxford Studies in Normative Ethics, Volume 9* (pp. 271–294). Oxford University Press

Rottschaefer, W. A., & Martinsen, D. (1990). Really taking Darwin seriously: An alternative to Michael Ruse's Darwinian metaethics. *Biology and Philosophy, 5*, 149-173.

Ruiz-Mirazo, K., Peretó, J., & Moreno, A. (2004). A universal definition of life: autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere, 34*, 323-346.

Ruse, M. (1984). The morality of the gene. *The Monist, 67*(2), 167-199.

Saborido, C., & Moreno, A. (2015). Biological pathology from an organizational perspective. *Theoretical Medicine and Bioethics, 36*(1), 83-95.

Saborido, C., Mossio, M., & Moreno, A. (2011). Biological organization and cross-generation functions. *The British Journal for the Philosophy of Science, 62*, 583–606.

Saborido, C., A. Moreno, M.González-Moreno & J. C. Hernández Clemente (2016). Organizational Malfunctions and the Notions of Health and Disease (pp. 101-120). In E. Giroux (Ed.). *Naturalism in the Philosophy of Health* (History, Philosophy and Theory of the Life Sciences, 17). Dordrecht: Springer.

Sachs, C. B. (2019). In defense of picturing; Sellars’s philosophy of mind and cognitive neuroscience. *Phenomenology and the Cognitive Sciences, 18*, 669-689.

Savage, J. (2007). "Black magic" and white terror: slave poisoning and colonial society in early 19th century Martinique. *Journal of Social History, 40*(3), 635-662.

Schlosser, G. (1998). Self-re-production and functionality. *Synthese, 116*, 303-354.

Schmidtz, D. (1994). The institution of property. *Social Philosophy and Policy, 11*(2), 42-62.

Schmidtz, D. (2023). *Living together : inventing moral science*. Oxford University Press.

Schroeder, T. (2004). New Norms for Teleosemantics. In Clapin, H., Slezak, P., & Staines, P. (eds.) *Representation in Mind* (Vol. 1). Elsevier Science & Technology. https://doi.org/10.1016/B978-008044394-2/50008-7

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences, 106*(35), 15073-15078.

Sellars, W. (1962). Philosophy and the scientific image of man. *Frontiers of science and philosophy, 1*, 35-78.

Setiya, K. (2024). Human nature, history, and the limits of critique. *European Journal of Philosophy, 32*(1), 3-16.

Sfara, E., & El-Hani, C. N. (2023). Ecosystem health and malfunctions: an organisational perspective. *Biology & Philosophy, 38*(5), 37.

Shapiro, Lawrence and Shannon Spaulding. (2021) Embodied Cognition. In: Edward N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), URL = <https://plato.stanford.edu/archives/win2021/entries/embodied-cognition/>.

Shenk, M. K. (2024). Marriage. In: J. Koster, B. Scelza, & M. K. Shenk (Eds.), *Human Behavioral Ecology* (pp. 230–255). Cambridge University Press.

Sidgwick, H. (1902). *Lectures on the Ethics of T. H. Green, Mr Herbert Spencer, and J. Martineau* (E. E. C. Jones, Ed.). Cambridge University Press

Singer, P. (1981). *The expanding circle : Ethics and sociobiology*. Oxford: Clarendon Press.

Sixtus, E., Krause, F., Lindemann, O., & Fischer, M. H. (2023). A sensorimotor perspective on numerical cognition. *Trends in Cognitive Sciences, 27*(4), 367-378.

Smith, D. L. (2021). *Making Monsters : The Uncanny Power of Dehumanization*. Harvard University Press.

Smith, M. (1994). *The Moral Problem*. Blackwell.

Spencer, H. (1879). *Data of ethics*. D. McKay.

Stevenson, C. L. (1945). *Ethics and language*. Yale University Press.

Stingl, M., & Collier, J. (2019). *Evolutionary Moral Realism* (1st ed.). Routledge.

Stocker, M. (1976). The schizophrenia of modern ethical theories. *The Journal of philosophy, 73*(14), 453-466.

Strauss, L. (1953). *Natural Right and History*. University of Chicago Press.

Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies*, *127*(1), 109–166.

Street, S. (2009). In defense of future Tuesday indifference: Ideally coherent eccentrics and the contingency of what matters. *Philosophical Issues, 19*, 273-298.

Street, S. (2012). Coming to Terms with Contingency: Humean Constructivism about Practical Reason. In: J. Lenman, and Y. Shemmer (eds), *Constructivism in Practical Philosophy*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199609833.003.0003

Streumer, B. (2017). *Unbelievable Errors*. Oxford: Oxford University Press.

Thompson, E. (2007). *Mind in life : biology, phenomenology, and the sciences of mind*. Harvard University Press.

Thompson, M. (2004). Apprehending human form. *Royal Institute of Philosophy Supplements, 54*, 47-74.

Thompson, M. (2008). *Life and action : elementary structures of practice and practical thought*. Harvard University Press.

Tolstoy, L. (1899). *What is Art? ; What is Religion*. Crowell.

Ullmann-Margalit, E. (2017). *Normal Rationality*. Oxford University Press.

Tomasello, M., & Vaish, A. (2013). Origins of Human Cooperation and Morality. *Annual Review of Psychology, 64*(1), 231-255.

Van Duijn, M., Keijzer, F., & Franken, D. (2006). Principles of Minimal Cognition. *Adaptive Behavior, 14*(2), 157-170.

Velleman, J. D. (2014). What Good Is a Will? In: M. Vargas & G. Yaffe (eds.) *Rational and Social Agency*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199794515.003.0005

Verhage, M. L., Schuengel, C., Madigan, S., Fearon, R. M., Oosterman, M., Cassibba, R., ... & Van IJzendoorn, M. H. (2016). Narrowing the transmission gap: A synthesis of three decades of research on intergenerational transmission of attachment. *Psychological Bulletin, 142*(4), 337.

Voorhoeve, A. (2003). The grammar of goodness: an interview with Philippa Foot. *Harvard Review of Philosophy, 11*, 32-44.

Walsh, D. (2022). Being Human, Being *Homo sapiens*. In: Hübner, K. (Ed.). *Human : a history*. Oxford University Press.

Watkins, M. (2013). A Cruel but Ancient Subjugation?: Understanding Hume's Attack on Slavery. *Hume Studies, 39*(1), 103-121.

Weber, A. & Varela, F.J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences, 1*, 97-125.

Wenseleers, T., Helanterä, H., Hart, A., & Ratnieks, F. L. (2004). Worker reproduction and policing in insect societies: an ESS analysis. *Journal of evolutionary biology, 17*(5), 1035-1047.

Wenseleers, T., & Ratnieks, F. L. (2004). Tragedy of the commons in *Melipona* bees. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 271(suppl_5),* S310-S312.

Whyman, T. (2016). Understanding Adorno on 'natural-history'. *International Journal of Philosophical Studies, 24*(4), 452-472.

Whyman, T. (2017). Adorno's Aristotle critique and ethical naturalism. *European Journal of Philosophy, 25*(4), 1208-1227.

Williams, B. (1981). Internal and external reasons. In *Moral Luck: Philosophical Papers 1973–1980* (pp. 101–113). chapter, Cambridge: Cambridge University Press.

Williams, B. (1985). *Ethics and the limits of philosophy*. Fontana.

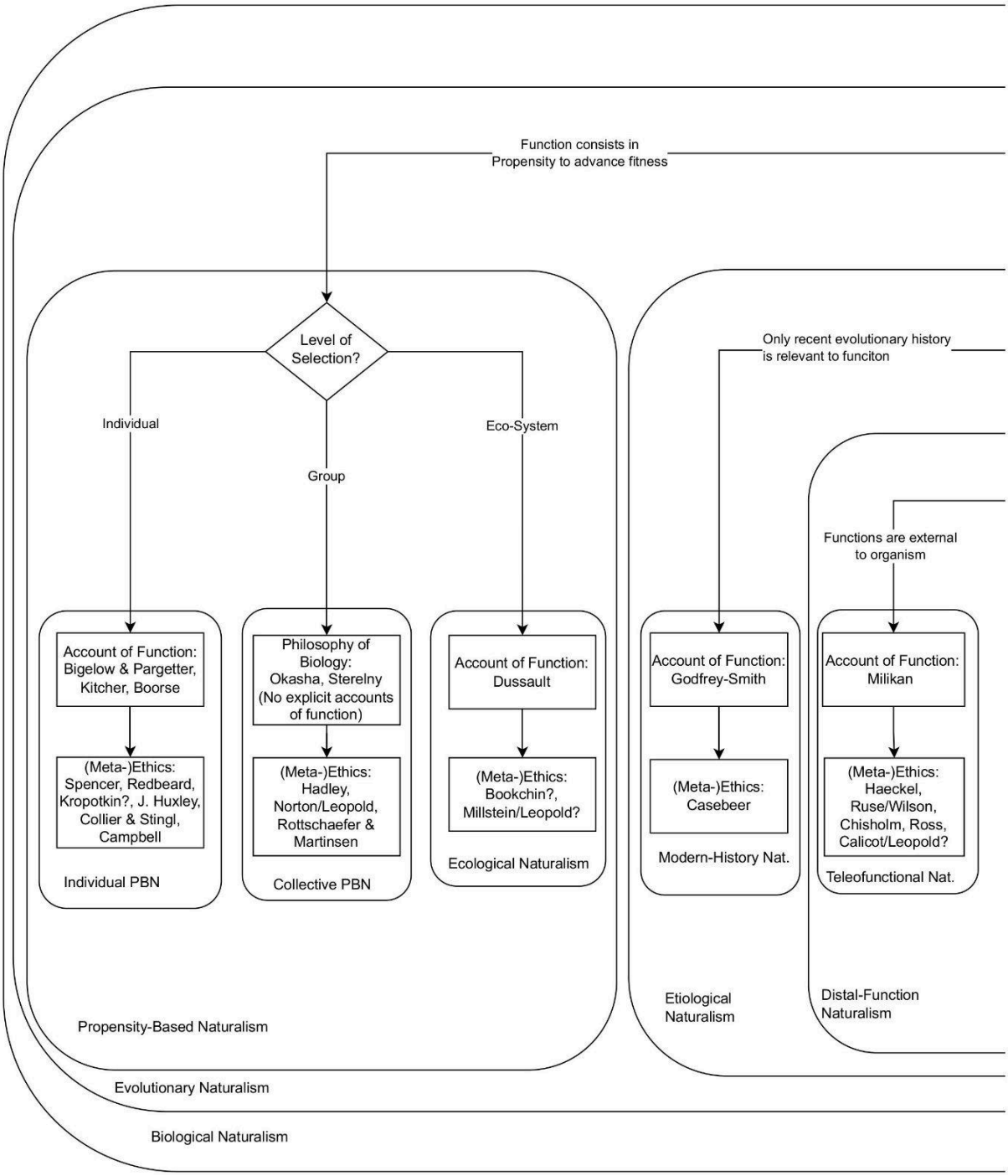
Williams, B. (1993). *Shame and necessity*. University of California Press.

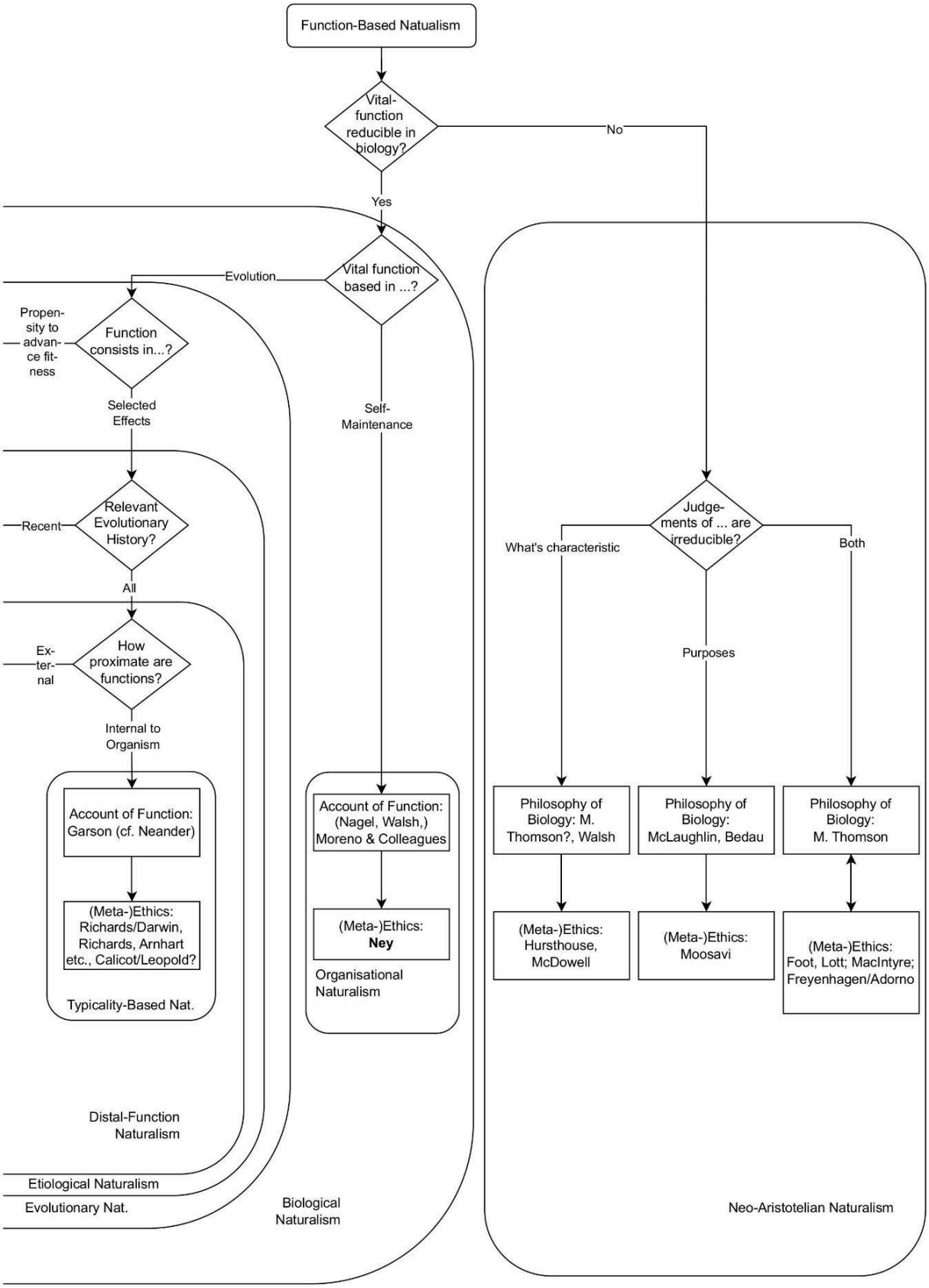
Wilson, J. Q. (1993). *The moral sense*. Free Press.

Wong, D. (2006). *Natural moralities : A defense of pluralistic relativism*. Oxford: Oxford University Press.

Woodcock, S. (2006). Philippa Foot's virtue ethics has an Achilles' heel. *Dialogue: Canadian Philosophical Review/Revue canadienne de philosophie, 45*(3), 445-468.

Appendix A: Function-based naturalisms (a diagram)




Function-Based Natualism
Vital-function reducible in biology?
No
Yes
Vital function based in ...?
Evolution
Propen-sity to advan-ce fit-ness
Function consists in...?
Selected Effects
Self-Maintenance
Recent
Relevant Evolutionary History?
All
Ex-ter-nal
How proximate are functions?
Internal to Organism
Account of Function: Garson (cf. Neander)
(Meta-)Ethics: Richards/Darwin, Richards, Arnhart etc., Calicot/Leopold?
Typicality-Based Nat.
Account of Function: (Nagel, Walsh,) Moreno & Colleagues
(Meta-)Ethics: Ney
Organisational Naturalism
Distal-Function Naturalism
Etiological Naturalism
Evolutionary Nat.
Biological Naturalism
Judge-ments of ... are irreducible?
What's characteristic
Purposes
Both
Philosophy of Biology: M. Thomson?, Walsh
Philosophy of Biology: McLaughlin, Bedau
Philosophy of Biology: M. Thomson
(Meta-)Ethics: Hursthouse, McDowell
(Meta-)Ethics: Moosavi
(Meta-)Ethics: Foot, Lott; MacIntyre; Freyenhagen/Adorno
Neo-Aristotelian Naturalism