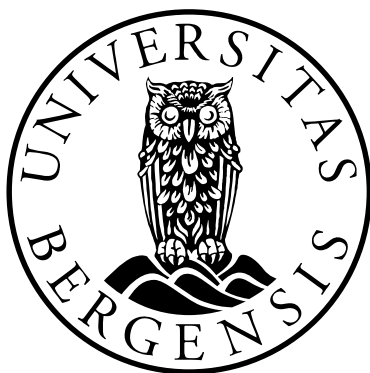


Hedonistic Egoism

A Theory of Normative Reasons for Action

Ivar Russøy Labukt



Dissertation for the degree philosophiae doctor (PhD)
at the University of Bergen

2010

Table of Contents

Preface	v
1. Introduction	1
1.1 Hedonism and egoism in the philosophical literature	1
1.2 Arguing for hedonistic egoism	4
1.3 The structure of the thesis.....	13
2. Reasons	15
2.1 Reasons for action	15
2.2 Reasons and other normative concepts.....	20
2.3 Reasons and desires	25
2.4 Ultimate reasons	31
3. Pleasure and Pain	33
3.1 The debate on pleasure and pain.....	33
3.2 The traditional hedonic tone view	35
3.3 The heterogeneity objection	38
3.4 Liking and motivation	41
3.5 The pluralistic hedonic tone view.....	44
4. Hedonism.....	53
4.1 Hedonism as a theory of ultimate reasons	53
4.2 The implications of hedonism	58
5. Egoism.....	68
5.1 Rational egoism	68
5.2 Attempts to prove that rational egoism is mistaken.....	76
5.3 Rational egoism and morality.....	84

6. The Implications of Hedonistic Egoism	95
6.1 The object of egoistic concern	95
6.2 The commitment to morality	97
6.3 The commitment to friendship	111
7. The case against ultimate reasons	118
7.1 Queerness.....	118
7.2 Disagreement	124
7.3 How there is an apparently plausible alternative.....	131
8. The case for hedonic egoistic ultimate reasons	142
8.1 Hedonic debunking of non-hedonic intuitions	142
8.2 Which intuitions survive hedonic debunking?	148
8.3 Can basic hedonic value judgments be debunked?	156
9. Objections to hedonism and egoism.....	169
9.1 Objections to the classical conception of hedonic egoistic reasons	169
9.2 Objections to hedonism	175
9.3 Nagel’s objections to egoism.....	181
9.4 Final remarks	190
Bibliography	192

Preface

My PhD thesis was originally supposed to defend hedonistic utilitarianism as a theory of morality. Rather coincidentally, the first part of this project I began working on was providing an answer to the common criticism that utilitarianism is too demanding, since it apparently requires that agents must dedicate most of their lives to the promotion of general happiness in order to act rightly. For some time I thought the problem could be solved by understanding utilitarian morality merely as one source of reasons for action among others: if there isn't necessarily most reason to do what is morally right to do, then utilitarianism would turn out not to be that demanding after all. I did, however, assume – as I had done for as long as I can remember – that there must at least be *some* non-derivative reason to make other people happy and to avoid making them miserable.

While trying to formulate my reply to the demandingness objection more precisely, I came across a paper that accused utilitarianism and other consequentialist moral theories of not being able to invest their moral principles with any authority at all – not even the quite modest one I had had in mind. The principle of utility might perhaps tell us what we ought to do from a moral or impartial point of view, but it is not clear, the author argued, why this is relevant at all when it comes to what each individual agent has reason to do, or what he ought to do in the all-things-considered sense. While I no longer think it was a particularly good paper, it got me thinking. Then, one day, I very suddenly found myself convinced that there *is* no non-derivative reason to be concerned with the interests of other people. I didn't want to reach this conclusion; I just did. The claim that one in some important normative sense just *has to* care about other people came to seem completely empty; there was simply nothing behind it. I even found it somewhat difficult to understand precisely what I had previously believed. I imagine many people have had similar experiences when they have given up their faith in God.

Changing my mind about the reason-generating power of other people's hedonic states did not undermine my conviction that I have a reason to promote my own pleasure and avoid my own pain. I had, in other words, come to believe in the hedonistic version of rational egoism. Several authors have defended the principle of utility as a criterion of moral rightness while at the same time suggesting that egoism may be correct as a theory of reasons for action, and so it would certainly have been possible for me to go through with my original project. However, it seemed much more tempting to try to articulate the basis for my newfound conviction that

there are no valid reasons for action except those that are grounded in one's own hedonic interests. After a few months of hesitation, I decided to give in to the temptation. The result is a thesis that is, I fear, less likely to meet with agreement, but at the same time, I hope, more interesting and original.

Many people have helped me write this thesis. First and foremost, I would like to thank my supervisor Erik Brown. I have lost count over how many drafts and dispositions I have presented him with, not to mention how many improvements and clarifications his thorough readings have enabled me to make. Also, I am grateful for his trust in letting me pursue my interests and arguments wherever they would lead me.

My co-supervisor Torbjörn Tännsjö provided several useful criticisms and suggestions, and was kind enough to invite me for a stay at the philosophy department at the University of Stockholm during the fall 2009. I had the opportunity to present material for my thesis at three seminars in Stockholm, and I received many helpful comments from the participants. In particular, I would like to thank Lisa Furberg for commenting on an early version of chapter 3, Jan Österberg for providing detailed and very useful comments on two earlier versions of chapters 5 and 6, and Ragnar Francén for reading a draft of the whole thesis and pointing out some important problems.

Thanks are also due to Roger Crisp, for commenting on a very early version of chapter 3, and to Ole Martin Moen (who is the only fellow believer I have met so far) for commenting on an earlier version of the whole thesis.

Between 2007 and 2009 I shared my office with Audun Stolpe. I greatly appreciate his moral support, and I enjoyed and benefited from our discussions about hedonism, egoism and many other philosophical topics.

Special thanks go to Hanne, for her patience and loyalty (and all the delicious late night dinners) during the hectic final year of writing, and to our son William, for making it even harder to think that rational egoism is true.

Finally, I would like to thank my mother, who has always supported and encouraged me, even when I made the somewhat unexpected and adventurous choice of studying philosophy. Although she most certainly did not raise me to be a hedonistic egoist, I dedicate this thesis to her.

1. Introduction

This thesis is an attempt to answer the question: “What does one have most reason to do?”

This is, as I understand it, the most basic normative question one can ask. It is equivalent to the question: “What ought one, ultimately and all things considered, to do?” If we agree that a given agent has most reason to do X, there is no room for disagreement on whether the agent is justified in doing X. Asking why one should do what one has most reason to do makes no sense.

The question that historically speaking has dominated normative philosophy is “What is morally right and wrong?” It is not always clear what this question means. Some authors define “morally right” as “that which there is most reason to do”, and then it is, of course, equivalent to the question I am discussing in this thesis. Others think that moral reasons are one class of reasons among others. If so, we need to know what other reasons there are and how strong they are compared to moral reasons. A moral theory would be a part of the complete theory of reasons. There are also philosophers who think that morality is not an independent source of reasons at all. On this view, we need a theory of reasons in order to show that we have any reason to be moral at all.

I will discuss the concepts of reason and morality in much more detail in later chapters. For now I just want to point out that a theory of reasons is the most basic normative theory there can be.

The theory of reasons I will be defending is hedonistic egoism. My claim will thus be that there is only one class of valid reasons for action: reasons grounded in one’s own future pleasure and pain. Each person has a reason to act in a way that gives him the most pleasurable life possible, and that is all there is to it; the idea that there are other reasons is ultimately just an illusion. In this introductory chapter, I will give an overview of my argument for this claim. First, however, I will say a few words about what other philosophers have thought and said about hedonism and egoism.

1.1 Hedonism and egoism in the philosophical literature

Hedonism and egoism have rarely been defended together. However, each individual theory has, of course, been the subject of much philosophical debate.

The history of hedonism is, at least in its main outlines, quite straightforward: it enjoyed some moderate popularity during antiquity, became the default position in 18th and 19th

century British moral philosophy, and then gradually lost its appeal among philosophers, to the point that even a careful and reasonable author such as John Rawls would characterize it as a “desperate expedient” and state that “surely the preference for a certain attribute of feeling or sensation above all else is as unbalanced and inhuman as an overriding desire to maximize one’s power over others or one’s material wealth” (Rawls 1999a, 488, 490).¹ Though hedonism does seem to have made something of a comeback the last few years (see Tännsjö 1998, Feldman 2004, Layard 2005, Crisp 2006, Mendola 2006, Brax 2009), it is still very much a minority position.

Egoism has a more complicated and disparate history. A form of the doctrine seems to have generally been taken for granted in Greek philosophy, and there is an egoist strain in much of contemporary virtue ethics (see Annas 2008 for a discussion). However, it must be kept in mind that the notion of eudaimonia in important respects differs from dominant modern notions of self-interest (see Österberg 1988, 1.2). The first modern egoist is Thomas Hobbes (see Hobbes 1651, esp. ch. XIV-XV). Some Christian writers, such as Samuel Clarke and Joseph Butler, also recommended the pursuit of self-interest, but since they believed that God will reward virtue in the afterlife, they thought there was no conflict between self-interest and virtue. It is not clear whether they would stick to their recommendation if God is removed from the picture (see Butler 1726; Österberg 1988, 1.3-1.6; Shaver 1999, 4.2-4.4).

Hume at times seems to treat considerations of self-interest as the ultimate practical standard (Hume 1743, 92; Hume 1751, 79-82), but it is unclear what his considered view is (see Shaver 1999, 4.5). It may also at first sight be tempting to treat Nietzsche as an egoist, but he is probably better understood as an opponent of traditional forms of altruistic morality than as an advocate of the exclusive pursuit of self-interest (see Nietzsche 1886, sect. 260, 265; Nietzsche 1887, II, sect. 18; Leiter 2002, 134-5; Risse 2008). Another 19th century German philosopher with a somewhat similar philosophical outlook, namely Max Stirner, was, on the other hand, an outspoken egoist (Stirner 1845). However, his version of the doctrine is very polemic, unsystematic and extreme, and appears to be of little contemporary relevance. Sidgwick’s treatment of egoism is, as one might expect, clearer and more sober (see Sidgwick 1907). Though he did not embrace the theory, he found it to be just as plausible as utilitarianism (and famously dubbed this standoff the “dualism of practical reason”).

¹ The choice of comparisons may suggest that Rawls is talking about the egoistic version of hedonism, but from the context it appears that his comments are directed towards hedonism in general.

In the 20th century, egoism was often taken to be a candidate for an ethical or moral theory, on a par with utilitarianism, Kantianism and the rest (see e.g. Brunton 1956; Medlin 1957; Kalin 1970; Österberg 1988). (Somewhat curiously, most of the discussion centered on the question of whether ethical egoism is consistent, and not whether it is substantively plausible.) David Gauthier's moral theory, which is similar to that of Hobbes, is probably the most sophisticated version of ethical egoism in the literature (see Gauthier 1986). His moral principles are not directly egoistic: they sometimes require genuine sacrifices of the agent's best interests. However, they are nevertheless, or so he argues, the kind of principles that someone trying to maximize his own utility would want to be disposed to follow: the idea is to choose "on utility-maximizing grounds, not to make further choices on those grounds." (Gauthier 1986, 158) Thus, moral principles, on Gauthier's view, ultimately derive their content and authority from egoistic principles.²

As I will argue in chapter 5, ethical egoism is either an implausible theory or a misleading description of rational egoism, i.e. egoism as a theory of reasons for action. The focus on ethical egoism has been unfortunate, since it has made philosophers pay less attention to rational egoism, or even overlook it completely. Now rational egoism does, of course, now and then make its appearance in the contemporary literature. Several prominent utilitarians have tried to show that even if rational egoism is true, we still have strong reasons to act in accordance with the principle of utility, since being disposed as a (reasonably) good utilitarian is in our best long-term interest (Hare 1981, ch. 11; Brink 1989, 8.6, Singer 1993, ch. 12). These authors do not exactly go out of their way to embrace rational egoism, but they appear to think that it might well be true, or at least that it is a kind of worst-case scenario they do not know how to refute.

Most often, perhaps, rational egoism figures in philosophical discussion as a background view that the author is trying to avoid or disprove (see e.g. Nagel 1986, 156-62; Korsgaard 1996, ch. 4). The egoist position is then usually not defined or explicated with any care, and the result is, as we will see examples of later in this thesis, that the anti-egoist arguments tend to miss their target. Of course there are also good and thorough accounts of the egoist position, such as those found in Sidgwick's *Methods of Ethics* and Parfit's *Reasons and Persons* (Sidgwick 1907, book II; Parfit 1984, part 1). However, I do not think that

² Ayn Rand is probably the defender of ethical egoism who has been most influential outside the academic world, but philosophers have generally found her writings overly ideological and rhetorical. An exception is Tara Smith, who has written a book on Rand's egoism (T. Smith 2006; for Rand's own statement of her view, see Rand 1964).

philosophers have spent enough time on developing a clear and maximally plausible version of rational egoism. Rational egoism is, after all, a very natural view on a very basic question.

There is also a striking absence of positive arguments for rational egoism in the literature.³ Hobbes seems to have simply taken it for granted (see Shaver 1999, ch. 1). Most other egoists seem to think that no positive argument is necessary: egoism is, as Robert Shaver puts it, conceived of as “the “default view” that any rival normative theory must defeat.” (Shaver 1999, 1) Gauthier thus offers no other defense of his egoistic starting-point than an assertion that the “onus of proof” is on the non-egoist (Gauthier 1986, 8), and Sidgwick’s insistence on the plausibility of egoism merely amounts to admitting that he cannot refute it (Sidgwick 1907, xx, 497-8, 507-9). It is questionable whether rational egoism deserves the status of a default view that requires no positive argument. In any case, it would of course be of significant philosophical interest if such an argument could be provided.

It is difficult to assess how plausible contemporary philosophers find rational egoism. In *Reasons and Persons*, Derek Parfit treats the theory as a strong contender and claims that it “has been believed by most people for more than two millennia” (Parfit 1984, 194). On the other hand, Roger Crisp, in his recent book on reasons, finds it sufficient to simply assert that “egoism is mistaken” and that those who disagree are “in the grip” of their theory “to the point where they cannot appreciate what to most of us is obvious.” (Crisp 2006, 132; see also Shafer-Landau 2003, 196-9; Parfit forthcoming, 143-4) Judging from my (relatively modest) personal experience, most philosophers are indeed quite skeptical of egoism, and some are so skeptical that they seem unwilling to give the theory a fair chance. At first, I thought this hostility was due to egoism being seen as an evil or pernicious view. However, it seems as if quite a few philosophers are also intellectually prejudiced against egoism: they think the view must be naïve, ideological or based on obviously fallacious reasoning. It may be that some versions of egoism really do meet this description. I hope this thesis will illustrate that not all of them do.

1.2 Arguing for hedonistic egoism

In this section, I present my argument for hedonistic egoism. It should be pointed out from the outset that the hedonistic and the egoistic aspects of the theory are, logically speaking,

³ For the sake of completeness, it should be mentioned that Ayn Rand tries to show that egoism follows from the fact that we are living organisms (see e.g. Rand 1964, ch. 1). A criticism of the argument can be found in Österberg 1988, 4.3.

independent: neither one follows from the other. One can be a hedonist and reject that one's own pleasure and pain is all that matters and one can be an egoist without accepting a hedonistic theory of self-interest. I have defended them together because they both seem to be supported by the considerations outlined below. If they aren't, or if the considerations are insufficient to establish any conclusions, then the reader may of course decide to accept one of these views and not the other.⁴

I will start by distinguishing my argument from certain other considerations that could be offered in favor of hedonism or egoism (or both).

Hedonistic egoism is an extremely simple and unified theory of reasons for action. It traces all normative authority back to a single source: the hedonic level of one's own experiences. In comparison, many other accounts only seem to provide us with more or less unconnected heaps of reasons. W. D. Ross, for instance, claims there are reasons to (among other things) keep promises, be grateful, promote knowledge and distribute pleasure according to moral worth (Ross 1930, ch. 2). There is no underlying normative principle explaining why these things provide us with reasons while many others are not. Authors such as Joseph Raz and Thomas Scanlon seem to think that the realm of reasons is even more extensive and diverse (see Scanlon 1998; Raz 1999).

Unity and simplicity are often considered theoretical virtues. Sidgwick puts great weight on the lack of simple and unified alternatives in his defense of hedonism (Sidgwick 1907, book III, ch. XIV, §3).⁵ David Brink also thinks there is, on any coherentist epistemology, a significant presumption in favor of unified normative theories (Brink 1989, 8.9). Now I do not want to deny that a high degree of unity and simplicity may, other things equal, to some extent increase the plausibility of a theory of ultimate reasons (see Baker 2010 for a general discussion of theoretical simplicity). However, if other things are not equal, the force of these considerations seems to be quite limited. Let's say that you believe that hedonistic egoism generally makes plausible claims about reasons, but that there are a few cases where it seems to go wrong. There are a number of legitimate reasons why you could decide that these

⁴ In *What We Owe to Each Other*, Thomas Scanlon argues that if we reject hedonism in favour of value pluralism, there is no basis for according paramount normative importance to the fulfilment of one's own interests (Scanlon 1998, 3.4). I find this argument reasonably persuasive, so it may be unlikely that egoism should turn out be correct if hedonism fails. On the other hand, I know of no general reason for thinking that a rejection of egoism would undermine the credibility of hedonism.

⁵ Perhaps this is why John Rawls thinks the whole hedonist tradition is driven by a need for a single, dominant end in all deliberation (Rawls 1999a, §84).

appearances probably are mistaken. For instance, you may find an explanation for the intuitions in question that undermines their credibility, or you may discover that making what seems to be plausible non-egoistic or non-hedonistic claims in these particular cases forces you into problems in other cases that are even more serious than the ones they were meant to solve. But let us assume that you have no such reasons: hedonistic egoism really does on reflection (apart from considerations of simplicity) appear to be substantively mistaken about the cases in question. Then it seems to me that the mere fact that hedonistic egoism is a simple and unified doctrine is not sufficient to show that it isn't mistaken after all. My defense of the theory will, at any rate, not rely on this kind of claim.

Hedonistic egoism may appear to be to some extent vindicated if it turned out that all human beings, as a matter of fact, already act in accordance with it. John Stuart Mill provided an infamous argument for hedonism of this type, based on the claim nothing except pleasure is ever desired for its own sake (Mill 1871, ch. 4). A similar argument could be given for rational egoism, and in fact many non-egoist philosophers seem to be under the impression that this is the standard way of arguing for the doctrine. However, it is not – and for good reason. The psychological versions of egoism and hedonism are both highly unlikely to be true, and in any case they do not seem to entail any normative views. Even if human beings were incapable of acting in a non-egoistic or non-egoistic manner, it could still be true that they would have reason to do so *if* they had been able to (see Österberg 1988, 4.2).

A more promising strategy, at least initially, is to try to show that all or most people hold normative *beliefs* that support hedonistic egoism. Of course there are few who explicitly endorse the view. But it could still be true that many or even most people would do so if they only understood the doctrine and its philosophical alternatives properly. Sidgwick thinks that common sense favors hedonism in this indirect way (Sidgwick 1907, book III, ch. XIV). In a more recent discussion, Joseph Mendola argues at some length that hedonism is more intuitively plausible than its competitors. The reason other philosophers have failed to see this is that they have either misunderstood the theory, failed to consider it in its most plausible variety, made mistakes when conducting thought experiments, relied too much on intuitions about far-fetched and bizarre cases, or overlooked counterintuitive features of other theories (Mendola 2006).

I do not want to deny that all these things happen. Hedonism is often dismissed *too quickly* on intuitive grounds. However, I do not think that all critics of hedonism are wrong about the intuitive plausibility of hedonism all of the time. The doctrine is in genuine conflict with common sense and the intuitions of most philosophers. Consider, for instance, Nozick's

experience machine objection (which I will discuss in much more detail in chapter 9). Surely there are a number of complicating factors which can (and sometimes does) lead philosophers to overestimate the extent to which hedonism is counterintuitive in this particular case. But that doesn't change the fact that it also reveals a deep and widely shared intuition that some sort of authenticity matters for its own sake. This intuition is still there even we adjust for all distorting factors when conducting the thought experiment.

Similar considerations apply, I believe, to the egoistic aspect of the theory. Hedonistic egoism cannot, in other words, be defended by according a default epistemological weight to common sense or prevalent normative intuitions. If we do this, we will find many intuitions that count in favor of the doctrine, but there will also be many – too many – that count against it.

One could, of course, be a hedonistic egoist on intuitive grounds without claiming that the intuitions in question are shared by all or most other people. If the theory appears intuitively plausible to you, why care about what others think? In this vein, it is sometimes said that rejections of egoism based on appeals to normative intuitions are unsatisfactory, since there is no guarantee that the egoist will share the intuitions in question (see e.g. Österberg 1988, 5.1). However, we should not expect or require unanimous support for the correct normative theory. There will always be people around with all sorts of crazy normative convictions. If the hedonistic egoist simply is a person with idiosyncratic intuitions, then those who have diverging views need not worry too much about the fact that there is no way of proving, in any strict sense, that he is mistaken.

In particular, we should be suspicious of people who seem to *like* egoism (or hedonism, for that matter). It is tempting to think that they do so because they are megalomaniacal, because they enjoy strife, because they have problems with authorities, or because they lack the capacity for sympathy and for deep and mutually rewarding interpersonal relationships. Max Stirner is an egoist of this kind. "I would", he says, "rather be referred to men's selfishness than to their "kindness", their mercy, pity, etc. [...] The affectionate one's service can be had only by—*begging*, be it by my lamentable appearance, by my need of help, my misery, my—*suffering*." (Stirner 1845, 310) He ends his book *The Ego and his Own* by a triumphant declaration of self-sufficiency:

I am the owner of my might, and I am so when I know myself as unique. [...] Every higher essence above me, be it God, be it man, weakens the feeling of my uniqueness, and pales only before the sun of this consciousness. If I

concern myself for myself, the unique one, then my concern rests on its transitory, mortal creator, who consumes itself, and I may say:

All things are nothing to me. (Stirner 1845, 366)

My defense of hedonistic egoism is not motivated in this way. I do not want the view to be true; I just find that there are good reasons to think that it probably is.

What are these reasons, then? The rough idea is that hedonistic egoism is supported by certain considerations that are typically offered in favor of normative anti-realism. Anti-realists believe that, as seen from a philosophical point of view, the world contains no values or reasons for action. Usually, they do not deny that there *appears* to be values or reasons for action in the realist sense.⁶ However, they think that the appearances are illusory: reflection shows that there is ultimately nothing normative behind them.⁷ Realists, on the other hand, think that there is. Unfortunately, there is a tendency in the debate to treat the realist appearances as a fairly undifferentiated class: either they are all misleading and anti-realism is correct, or they are a generally reliable source of knowledge of realist normative truths. Thus, most contemporary moral philosophers will fall into one of two categories: they are either anti-realists who think there is no normative reality at all or realists who think that the normative reality is roughly as it appears to be (as seen from a contemporary Western point of view, that is).

There is, of course, a third alternative.⁸ Many or even most normative appearances may be illusory, but perhaps a few of them do withstand philosophical scrutiny. Though I am sure that no philosophers would deny that this could conceivably be the case, there aren't many who actually defend views of this kind. This is, however, what I will do in this thesis. I will argue that the intuition that there is a reason to seek one's own pleasure and to avoid one's own pain

⁶ Of course some anti-realists do deny this. I will not discuss their views in any detail in my thesis, but instead simply appeal to the relatively wide consensus that they are mistaken. At least in a significant number of cases, what appears to be valuable also appears to be (in some sense) objectively valuable.

⁷ It should be noted that relativists typically think there is a genuine normative reality; it is just that it is not objective in the way a realist thinks it is. I give a brief criticism of certain forms of relativism about reasons in section 2.1.

⁸ There is, in fact, also a fourth possibility: one could try to justify a realist position without relying on normative intuitions at all. Kantians sometimes take this approach (see e.g. Wood 2008, ch. 3). I criticize some of these arguments in later chapters.

is not vulnerable to debunking of the kind attempted by anti-realists, while all other (substantive) normative intuitions are.

It is important not to confuse the intuition I will defend as uniquely trustworthy with an intuition that *only* one's own hedonic states ground ultimate reasons, i.e. that hedonistic egoism is true. I do not have this latter intuition (and I would not accord it any credibility if I did). In fact, I still have lots of non-egoistic and non-hedonistic intuitions.⁹ My conviction that nothing *besides* my own pleasure and pain matters for its own sake is entirely a result of philosophical reflection.

Why single out the agent-relative hedonic intuition for special treatment? The full answer will be given in chapter 7 and 8. Let me, however, give a brief summary. Anti-realism about reasons is only plausible if there is some explanation of why there sometimes appears to be reasons in the realist sense. Traditionally, anti-realists have embraced some form of projectivism: we believe in realist reasons because our subjective responses “color” the things we encounter, so that the negativity or positivity of the response is mistakenly experienced as belonging to the things themselves. Recently, a number of findings from psychology and neuroscience have been taken to support this account of the genesis of intuitions about reasons and values. I will argue that philosophers who draw a general projectivist conclusion from these findings overlook the crucial role hedonic states play in the formation of intuitive normative beliefs. It is these states that give our subjective responses their positive or negative character, and so without them the kind of normative projection anti-realists talk about wouldn't occur in the first place. So while the thesis that non-hedonistic and non-egoistic intuitions are merely the results projection is plausible and to some extent empirically supported, this does not give us any reason to think that the same goes for our immediate judgments concerning the value of our own hedonic experiences.

Even if these immediate judgments cannot be undermined by a demonstration of their hedonic origin, there may of course be other ways to debunk them. I will consider the most important ones, arguing that they are all implausible, or at least *much* more speculative and in need of further support than what anti-realists take their general projectivist thesis to be. Thus,

⁹ Walter Sinnott-Armstrong defines a normative intuition as a non-derivative and immediate normative appearance that is *endorsed* (Sinnott-Armstrong 2008, 47-8). In other words, you cannot have an intuition without having an intuitive belief. If this is how we understand intuitions, I could not say that I still have non-egoistic and non-hedonistic intuitions. I would only have experiences of apparent non-egoistic and non-hedonistic reasons for action. For the sake of simplicity, I will use the term “intuition” in a broader sense than Sinnott-Armstrong, so that it includes appearances believed to be mistaken.

the upshot will be that it is not arbitrary but quite reasonable to think that the intuitions that there is a reason to seek one's own pleasure and avoid one's own pain constitute the only exceptions to the projectivist thesis. This would in turn, of course, mean that hedonistic egoism is the correct theory of reasons for action.

This form of argument is not entirely foreign to the hedonist tradition. In *Utilitarianism*, John Stuart Mill tried to explain away non-hedonistic convictions by claiming that people have desires for non-hedonic objects because these objects have previously been associated with pleasure (Mill 1871, ch. 4). Later hedonists have often followed in his footsteps (see e.g. Silverstein 2000). There are two problems with Mill's account. First, it is speculative and based on a rather simplistic associationist psychology. Second, it aims to explain why people have non-hedonic *desires*, but objections to hedonism typically start from intuitive *judgments*. My argument in this thesis avoids both these problems, since it is based on extensive empirical research concerning precisely intuition and judgment. This research has apparently been generally overlooked by hedonists, which is quite surprising, given how it offers a fairly obvious possibility for strengthening their view.¹⁰

When hedonistic egoism is defended on these grounds, it is easy for a supporter of the theory to explain why most philosophers have failed to grasp the truth of this apparently simple doctrine: they have been systematically misled by hedonic projection. The hedonistic egoist can, in other words, explain away the intuitions that conflict with the theory without undermining the credibility of the intuitions that *support* the doctrine. On the other hand, if the normative reality is supposed to be roughly as it appears to be, then it is much more difficult to understand why there is so little convergence in normative philosophy. In particular, those who want to defend pluralist theories of reasons have a hard time explaining why it is that their own intuitions should be considered reliable while those of other pluralists with diverging views should not.

¹⁰ It should be noted that Leonard Katz, in a brief paper, claims that the research in question supports hedonism rather than anti-realism (Katz 2008). He does not argue for this any detail, though. As far as I know, the only author to use the findings in a proper argument for hedonism is David Brax, in his recent PhD thesis *Hedonism as the Explanation of Value* (Brax 2009). (He does not, however, say much about the anti-realist alternatives.) In one sense, his aim is less ambitious than mine, since he only defends axiological hedonism and makes no claims about the content of hedonic reasons for action. In another sense, however, it is much more ambitious, since he does not just argue that pleasure is the only thing that has value, but that pleasure *is* value, and, moreover, that this kind of value can be understood in purely naturalistic terms. My version of hedonism is not intended to be naturalistic, and I briefly criticize a naturalistic conception of hedonic value in section 7.1. However, I do believe that if I am wrong about the merits of ethical naturalism, Brax' view would be highly plausible.

Thus, I believe that anti-realists have been too hasty in taking the amount of persistent normative disagreement and recent developments in empirical moral psychology to undermine normative realism in general. What is undermined is only the idea that many or most of our intuitions are reflections of realist truths. When we get rid of this assumption, we open the door to a new and promising way to vindicate both realism about reasons for action and a particular first-order theory of such reasons.

My defense of hedonistic egoism involves, then, a generally critical attitude towards normative intuitions. This attitude should not be taken to express any particular methodology or conviction about how philosophy ought to be done. It is not as if I approached the normative realm with a firm belief that intuitions should be disregarded, and then proceeded to try to make sense of the normative realm on the basis of this belief. Rather, I started with trying to make sense of the normative realm, and found that the best way to do so was to consider a large class of intuitions to be mistaken. My skepticism towards these intuitions has to be defended by argument in the same way as all my other claims. It is a part of the package I am offering, and not the basis upon which the package was put together.

It might perhaps seem as if the claim that normative intuitions are systematically misleading is incompatible with Rawls' widely accepted suggestion that we should aim for a reflective equilibrium in normative philosophy: a correspondence between general principles and considered judgments about particular cases. We must, however, distinguish between narrow and wide reflective equilibria. In order to reach a narrow equilibrium, you simply look for the principles that best match your own pre-theoretical convictions. Presumably you will not be able to develop principles that cohere perfectly with all your particular judgments, so you will have to adjust these judgments somewhat. But it remains true that the only basis for the narrow reflective equilibrium is the convictions you start out with. The instruction to seek a narrow equilibrium would be a clear ethical methodology: a procedure that more or less automatically leads you towards a justified set of general principles. Unfortunately, it would also be a very unreasonable methodology. As several critics have pointed out (see e.g. Brandt 1979, 16-22; Hare 1989, 146), it is unacceptably conservative; basically, it simply amounts to a systematization of one's prejudices. Depending on our starting points, we will reach widely different conclusions about what we have reason to do.

Rawls made it clear, however, that what he is ultimately after is not a narrow equilibrium, but a wide one (Rawls 1999a, 43, Rawls 2001, 30-1). A wide reflective equilibrium includes knowledge of the alternative positions and the arguments that can be provided for them. In

other words, the principles must withstand the criticisms made by other philosophers with other convictions. They must also, as Norman Daniels has stressed, be compatible with (or preferably favoured by) certain background facts, such as a conception of the person or general psychological and sociological theory (Daniels 1979). This understanding of the notion of reflective equilibrium makes it possible to avoid the charge of conservatism and relativism. The cost, however, is that we are left with something that is no longer a distinct methodology. What Rawls instructs us to do is basically just to choose the principles that on reflection seem most plausible, remembering to consider all available alternatives and take all relevant convictions into account (Scanlon 2003; see Rawls 2001, 133-4; Rawls 1999b, 289).¹¹

In particular, the notion of a wide equilibrium is neutral between normative conservatism and revisionism. As Rawls says, when we search for this kind of equilibrium, our views “may or may not undergo a radical shift” (Rawls 1999a, 42). There is no way of knowing a priori how far the revisions will go: it all depends on where the arguments take us. Even Richard Joyce, who holds that morality is a complete illusion generated by certain evolutionary and psychosocial mechanisms, insists that his view is capable of sustaining a wide reflective equilibrium (Joyce 2001, 6.6).

In *What We Owe to Each Other*, Thomas Scanlon gives an account that is similar to Rawls’. He suggests that we start with the things that seem to be reasons, but he acknowledges that the appearances may be “illusory” and he puts no restrictions on how often this could turn out to be the case (Scanlon 1998, 1.12). In fact, my approach is even compatible with W. D. Ross’ so-called classical intuitionism, at least at it is understood and defended by Robert Audi. Audi thinks any particular intuitive judgment is both fallible and defeasible, so that in order to decide which intuitions to trust, we have to go through the kind of reflection involved in seeking a wide reflective equilibrium (Audi 2007, I.2).

Rawls, Scanlon and Audi all end up trusting many more intuitions than I do. As I hope to have shown, this is not because we are doing different kinds of normative philosophy. We simply disagree on a substantive question.

As is perhaps obvious by now, this thesis discusses both first-order and second-order or metaethical claims. These kinds of claims are, of course, logically separable. However, when

¹¹ What would offend against the “method” is to believe in a general principle and at the same time make a particular judgment that conflicts with the principle. But presumably no moral philosopher is under the illusion that doing so is methodologically sound.

it comes to arguing for their plausibility, I have found that it is best not to make any sharp distinction between normative and metaethical issues. In this sense, my thesis adopts the approach to practical philosophy shared by late 19th and early 20th century authors such as Henry Sidgwick, G. E. Moore and W. D. Ross, rather than the narrower focus on purely normative or purely metaethical questions found in much of contemporary philosophy. Again, this is not the result of some a priori conviction about the nature of normative philosophy. It just seems to me that there is a way of resolving some of the first- and second-order debates at the same time. If my argument is unsuccessful, others can approach these questions however they see fit.

Finally, it should be noted that defending a more or less complete first- and second-order theory of reasons is a very ambitious and comprehensive project. I cannot address all the questions that are relevant for this task, and when it comes to those I do address I cannot give all of them all the attention they deserve. I hope that what this thesis lacks in detail and thoroughness is compensated by the significance of its central claims.

1.3 The structure of the thesis

I end this introduction by providing a brief overview of the chapters to come.

In chapter 2, I clarify the concept of a reason for action. Mostly, my aim will not be to establish any substantial claims about what reasons there are, but only to lay the terminological foundations for the following discussion. I will, however, criticize the view that reasons are grounded in desires or preferences.

In chapters 3-6, I explain the content of hedonistic egoism as a theory of reasons. I start with the hedonistic aspect of the theory. In chapter 3, I discuss the nature of pleasure and pain and argue that these states are distinguished by their phenomenal feel, or hedonic tone. I then turn to hedonism and its implications in chapter 4. Chapter 5 starts with a description of rational egoism. After criticizing some attempts to show that rational egoism is inconsistent or based upon unacceptable presuppositions, I discuss how accepting rational egoism would affect our beliefs about morality.

In chapter 6, I give an account of what I take to be the practical implications of rational egoism, on a hedonistic theory of self-interest. I will argue that hedonistic egoism – contrary to popular opinion, perhaps – is not a particularly revisionist doctrine when it comes to the conduct it recommends in ordinary life cases. The conflict between hedonistic egoism and common sense is found in the way this conduct is justified (it is also revealed in various thought-experiments). With a few exceptions, the doctrine implies that most Westerners

should lead their lives in roughly the same way as they do today. Seeking non-hedonic goods like knowledge or beauty and being nice to one's friends and people in general typically improves the hedonic level of our experiences, at least in a long-term perspective. I defend this claim because it appears to be true, and not because I want to make hedonistic egoism seem more innocuous than it is. The claim plays no direct role in my argument for the theory.

The last three chapters contain my argument for hedonistic egoism. In chapter 7, I present some well-known objections to normative realism and try to show that several of the responses in the literature on behalf of realism are unsatisfactory. I then go on to argue, in chapter 8, that these objections have much less force when it comes to reasons that are based on one's own pleasure and pain. This does not, by itself, settle the question of whether these reasons are valid in the realist sense; it only shows that the case must be judged on its own merits. I discuss both realist and anti-realist understandings of egoistic hedonic reasons and argue that the realist alternative is, in the end, more plausible.

The discussion in chapter 7 and 8 will remain quite general. In chapter 9, I illustrate how my outlook deals with more specific objections that have been made to hedonism and egoism.

2. Reasons

In this chapter, I clarify the notion of a reason for action and explain how it is related to other normative concepts. It should be pointed out from the outset that my terminology is not meant to follow ordinary philosophical usage as closely as possible. (Many of the terms in question have no clear established meanings in any case.) Nor is it proposed as a revisionary vocabulary that anyone who wants to offer a theory of reasons for action ideally ought to adopt. Other substantive views may require other terminological choices. I have chosen the terms that allow the clearest and most convenient expression of my own claims.

2.1 Reasons for action

This thesis is about reasons for action, and not reasons for belief. It is important to note that these two kinds of reasons do not strictly speaking compete or conflict; they are simply reasons for different things (see Crisp 2006, 36-7). To say that something is a reason for believing a proposition is to say that it constitutes evidence for the proposition being true. It is the task of epistemology to specify what counts as evidence for a proposition being true. Whatever the correct epistemological view is, it has no direct implications for what we have reason to do.

Some philosophers think that it may be possible to choose what to believe: the formation of a belief could itself count as a voluntary action (see Raz 1999, ch.1 for a discussion). We could then ask whether the fact that a belief is epistemologically justified entails that there is a reason to choose to form it. The answer is that it does not. It is perfectly possible that there is no reason to choose to form a justified belief. Imagine, for instance, that you are shown convincing evidence for the proposition that Paraguay is a larger country than Uruguay. If you do not care about geographical facts, you may not have any reason to perform the action of forming this belief. It is also possible that you do: for instance, W. D. Ross thought we have reason to seek all kinds of factual knowledge for its own sake (Ross 1930, 139-49). The important point is that no epistemological theory can decide this question. It must be answered by practical philosophy or, more precisely, by a theory of reasons for action.

Other philosophers think that beliefs are formed involuntarily and automatically. There could then be no reasons to choose to form beliefs. However, we could still ask whether there is any reason to perform actions that indirectly will lead you to have epistemologically justified beliefs, such as searching for evidence or, once a belief is formed, spending some

mental effort rehearsing it, so that you won't forget it. Again, epistemology provides no guarantee that there are such reasons. Even if beliefs about the relative size of countries in South America can be epistemologically justified, one need not have any reason to seek out the evidence for them or strive to retain them.

It might seem as if there would at least have to be a reason of the kind that is relevant for action for deciding to believe or trying to discover the correct theory of reasons for action. However, whether there is such a reason depends on what the correct theory of reasons is like. If hedonistic egoism is true, you will have a reason to become a hedonistic egoist only if doing so is the best way of promoting your future hedonic level. For many people, it presumably would be. However, it need not always have this effect. Perhaps the theory makes you feel you are disillusioned and miserable, and that you would end up with more pleasure by believing some other theory, such as for instance Kantianism. If there is anything you can do in order to get yourself to believe in Kantianism, you would then have a reason to do it. This would, of course, not make the belief in Kantianism epistemically justified. Neither would it show that hedonistic egoism is not the correct theory of reasons for action after all (see Parfit 1984, 17-24, 43-5).

As has often been noted, reasons can be invoked both to explain and to justify actions (see e.g. Shafer-Landau 2003, p. 178-9; M. Smith 1994, 4.2; Crisp 2006, 38-9). These are, at least on the face of it, distinct activities. For instance, we may cite someone's envy as the reason for his actions, meaning that the envy makes his behavior understandable or a fitting object of an intentional stance. However, this does not mean that envy ever justifies any actions. The theory I defend in this thesis is intended as an account of justifying reasons. The idea is that conduct can be justified only by its consequences for the hedonic level of one's experiences, not that human beings are never motivated by anything other than their own pleasure and pain.

Many justifying reasons are instrumental or derivative: we only have a reason to do something because we have a reason to do (or bring about) something else. From the point of view of normative philosophy, the most interesting reasons are those that are non-instrumental or non-derivative. Now we sometimes talk about non-instrumental reasons for action that have no independent authority or normative force. Consider reasons of etiquette, for instance. There is, it seems natural to say, an etiquette-reason not to pick your nose independently of whether doing so will further any non-etiquette aims. However, most of us would agree that this reason lacks independent authority: it is not *by itself* relevant for practical deliberation. One could insist that a claim to authority is implicit in the notion of a reason for action, so that

a reason without such authority is not really a reason at all. There would then, presumably, be no non-instrumental reasons of etiquette. I find it convenient, however, to be able to talk about reasons even where there is no (independent) authority to be found, and I will do so every now and then throughout my thesis. Nothing substantial depends on this terminological choice.

The theory defended in this thesis is about authoritative reasons. The authority I am interested in is not just psychological authority. In order to have an authoritative reason to do something, it is not enough that one feels compelled to do it. People feel compelled to do all sorts of things that they do not have any reason to do. The authority is supposed to be of a normative kind. Perhaps the best way of explicating this notion of authority is to say that it makes no sense to ask why you should do what you have most authoritative reason to do. It is implicit in the idea of authority that this question cannot meaningfully be asked. You can only wonder whether there really are any authoritative reasons, or whether a given consideration provides you with one.

To say that a reason is authoritative is not to say that it is decisive. It may be outweighed or overridden – but only by other authoritative reasons. As I am using the concept, there can be no authoritative consideration in favor of an action other than a reason. Anything that authoritatively counts in favor of an action is ipso facto a reason for that action.

I do not think it is possible to give any further analysis of the notion of an authoritative reason. One can, of course, refer to it by other words. I have suggested that authoritative reasons have a particular normative force. When Jean Hampton tries to explain the nature of authoritative reasons, she says that they are prescriptive, binding and compelling, and that one in some sense just “has to” do what one has reason to do (Hampton 1998, 85-93). These expressions do not seem to add anything new, however. Thus, I am forced to take the notion of an authoritative reason as a basic idea and a starting-point of my inquiries in this thesis. I do not, however, think that it is a particularly controversial or unusual starting-point: most of the philosophers who are discussing reasons for action seem to accept something like it (see e.g. Crisp 2006, 8; Scanlon 1998, 3, 17; Parfit forthcoming, sect. 20) – though not necessarily under the name I have given it here (see e.g. B. Williams 1985, ch. 1; Tännsjö 2010, 3.8).¹²

¹² Some authors think that reasons can have different forms of authority: some reasons by their nature demand that a certain action be performed, while other reasons by their nature just make the action permissible or eligible (see e.g. Hampton 1998, 90-1; Raz 1999, ch. 5; Dancy 2004). On the version of hedonistic egoism defended in this thesis, there is no need for the latter kind of reason. Actions may of course be permitted rather than demanded by the doctrine, but that will only happen when two or more hedonic egoistic reasons with the same

Judgments about authoritative reasons are subject to what we might call a universalization requirement: by judging that an agent in a situation S has a reason R, one is committed to holding that any agent who is in a situation relevantly similar to S will also have the reason R. This requirement is not meant to exclude the view that reasons are subjective in the sense that they dependent on the agent's preferences or desires. It only excludes the claim (which probably no desire-theorist would make) that a desire could provide one agent with an authoritative reason for action, but fail to provide an authoritative reason for action if it is had by another agent in a situation that is identical in all relevant respects. To illustrate, the idea is that if I think that my desire to take a few day off work provides me with a reason to do so, I am committed to holding that any agent with a relevantly similar desire in an otherwise relevantly similar situation would have a reason to take a few days off. Claiming that the desire could be a source of an authoritative reason for me, but not for someone who is just like me and finds himself in identical circumstances, seems to make no sense. More generally, it is difficult to understand what someone using the expression "normative reasons for action" means if he denies the universalization requirement.

People make conflicting judgments about reasons for action. I defend hedonistic egoism, while other philosophers support some form of pluralism, and yet others think that all reasons are grounded in desires. We all accept the universalization requirement and apply our views to all agents. However, one can still wonder whether any of these views (or some other non-nihilistic view) is objectively correct. It seems to me that, if there are to be authoritative reasons for action at all, then this must be the case. Judgments about authoritative reasons for action are, in other words, subject to a further requirement: if an agent in a situation S has an authoritative reason R, this must be an objective fact. Let us call this the requirement of universal validity.¹³

As was the case with the universalization requirement, this requirement does not rule out the possibility that reasons are grounded in desires or other subjective features of an agent. The point is only that if a given subjective feature of an agent is supposed to provide the agent with an authoritative reason, then this must itself be an objective fact, and the corresponding judgment about what this particular agent has reason to do must be valid for all agents.

form of authority are equally strong. In the absence of competing hedonic egoistic reasons, any hedonic egoistic reason has the force to demand a given course of action.

¹³ Thomas Scanlon seems to accept both these requirements on reason judgments, but he lumps them together as a single "universality requirement" (Scanlon 1998, 73-4).

(Whether desires do in fact ground reasons for action is a question I will address in section 2.3.)

Furthermore, the requirement of universal validity is consistent with the doctrine that is sometimes referred to agent relativism (see e.g. Gowans 2010, sect. 2). According to agent relativism, what reasons an agent has depends crucially on what society or social group he is part of. For instance, we could imagine a form of agent relativism according to which Norwegians have most reason to pursue their own pleasure, while Swedes have most reason to promote human perfection. The judgment that Norwegians have most reason to seek their own pleasure would still be valid for all agents, including the Swedes, and so would the judgment that Swedes have most reason to promote human perfection. As long everyone can agree about what each Norwegian and Swede has most reason to do in this way, then there is no conflict with the requirement of universal validity. (There would of course have to be some relevant difference between the Norwegian and Swedish societies; otherwise the view would violate the requirement of universalization.)

The requirement of universal validity is, however, incompatible with so-called appraiser relativism, which is the view that different people can make conflicting reasons judgments *about the same agents*, without this entailing that any of them must be mistaken. Thus, we could, according to appraiser relativism, imagine that the Norwegians correctly judge that they (and everyone else, perhaps) have most reason to seek their own pleasure, while at the same time the Swedes correctly judge that the Norwegians (and everyone else, perhaps) have most reason to promote human perfection.

Appraiser relativism is a perfectly sensible doctrine if one thinks that the authority implicit in reason judgments is ultimately just psychological authority. It could certainly be true that Norwegians are most deeply committed to seek their own pleasure while Swedes are most deeply committed to promoting human perfection. In fact, Gilbert Harman, who is a prominent defender of appraiser relativism, seems to understand his doctrine in this fashion. After raising the question of whether relativist morality would have any normative authority, he says that “it is not so clear that such a question makes sense. Your standards of judgment derive from your superego, so, unless there is an internal conflict among those standards, it would seem that you must agree that forming and retaining the superego is desirable.” (Harman 1977, 62-3) A superego is, according to Harman, just a set of commands and prohibitions taken over from one’s parents, who in turn have derived them from their own parents, and so on (ibid., 60-2). Thus, it seems to be just a psychological mechanism. The fact that it would condone its own existence could make its psychological authority inescapable (at

least if we make the somewhat questionable assumption that all one's basic normative standards stem from one's superego), but it does not show that it has any authority that goes beyond the psychological.

One could also be an appraiser relativist about reasons judgments because one thinks that the term "reason" is used in different senses by different agents. Perhaps some are talking about instrumental reasons, others about motivating reasons and yet others about authoritative reasons. I have no doubt that variation of this kind occurs in everyday language. However, as philosophers keep their discussion fixed on a particular sense of the term "reason", this kind of relativism does not exclude the possibility that their judgments about reasons may have universal validity.

Finally, appraiser relativism could be intended as a doctrine about reasons with the kind of normative authority described above. This seems to me a very implausible view. I simply do not understand how it could be true for me that I am normatively compelled to seek my own pleasure if it is also true for other people that I am not normatively compelled to do so. It seems to me that either I am normatively compelled or I am not. Citing my failure to understand how appraisal relativism about authoritative reasons could be correct is perhaps not a very convincing objection to the doctrine. I will not, however, pursue the question any further. In the remainder of the thesis, I will disregard appraisal relativism as a possible solution to the problems I am discussing.

2.2 Reasons and other normative concepts

The views and arguments put forward in this thesis could all be formulated in terms of reasons for action. I will, however, also rely on some other normative concepts – partly to avoid repetition, but also because issues about reasons are sometimes discussed using these other concepts. In this section, I have a closer look at the concepts in question. As I have suggested, we should not expect this discussion to provide a reductive analysis of the concept of a reason for action. It will, however, clarify my terminology and help distinguishing the question I am interested in my thesis from certain other questions that have been discussed in the literature.

There must, obviously, be some connection between our reasons and what we ought to do. John Broome suggests that we should define a reason to do X as the *explanation* for why one ought to do X (Broome 2004; see also Tännsjö 2010, 3.1). I will settle for a simpler terminology: one ought to do X if and only if there is most reason to do X. Of course, the ought in question here cannot be the moral ought, unless one wants to assume from the outset

that morality is the only source of reasons for action.¹⁴ In order for the definition to work, it has to be what is sometimes called the practical ought, or the all-things-considered ought.

There is also a connection between reasons and *value*. The simplest account of this relation is the claim that being valuable or good is nothing over and above being a source of reasons for action. The things that are intrinsically good or bad would then be the things that are independent sources of authoritative reasons for action.¹⁵ Scanlon gives an analysis of value of this kind, which has come to be known as the buck-passing account (Scanlon 1998, 95-100). The buck-passing account is controversial (see e.g. J. Olson 2004; Crisp 2005), and in section 8.2 I will describe a possible scenario where it may break down. Nevertheless, I will, throughout my thesis, generally pretend that this outlook on value is correct, and take claims about value to entail claims about reasons for action. (I will also sometimes talk about things that matter for their own sake, and this is to be understood in the same way.) I do this because authors almost always believe that it is if not conceptually then at least substantively correct that what they take to be intrinsically good or bad provides people with reasons for action. It is hard to think of a philosopher who has argued that something is an important intrinsic value and at the same denied that it is in any way an independent source of reasons for action for any agent. Thus, when arguing about the kind of substantial matters I am interested in, it is a safe simplification to make no principled distinction between claims about reasons and claims about value.

This must not be taken to imply that I will assume that all reasons must be reasons to bring about as much value as possible. Perhaps some values demand not to be promoted, but to be, say, honored or protected. As Scanlon says, whether there are such non-teleological values is a substantive issue that should not be decided by terminological fiat (Scanlon 1998, ch. 2). In later chapters, I will argue that egoistic hedonic value is teleological: we have reason to maximize our own future hedonic level. However, when I talk about other possible values, the question of teleology is left open. I only assume that they are sources of at least some sort of reasons for action.

¹⁴ Some philosophers do assume this. I discuss different understandings of moral concepts in section 5.3.

¹⁵ Christine Korsgaard draws a distinction between intrinsic value, which something has in virtue of its intrinsic or non-relational properties, and final value, which something has if it is worth seeking or having for its own sake (Korsgaard 1983). My concern in this thesis is with what Korsgaard calls final value. However, I will argue that the things that are finally valuable, namely certain phenomenal feels, are so exclusively in virtue of their intrinsic properties, and it is therefore safe for my purposes to disregard Korsgaard's distinction. This is not to be taken as a denial that it may be important for the formulation of other substantial views.

If something is intrinsically valuable, to whom is it a source of reasons? We could of course stipulate that it would have to provide reasons for all agents. However, it would then be unclear whether anything actually has intrinsic value. It is better to leave the question open, and to only assume that intrinsic values are independent sources of reasons at least to some agent or other.¹⁶

One might think that saying that something is good *for* an agent, or a group of agents, would have to imply that it is a source of reasons to the agent(s) in question. Often, the expression “good for” is used in precisely this way. However, it is also used in discussions of well-being, and some authors evidently think that the vocabulary of well-being is distinct from that of other normative vocabularies. Roger Crisp, for instance, claims that it is an important substantial truth that an agent has reason to promote his own well-being, and then a further important substantial truth that hedonism is the correct account of well-being (Crisp 2006, ch. 3-4). I must confess that I find this notion of well-being difficult to understand. There does not seem to be two truths here, but just one: an agent has a reason to promote his own hedonic level. In any case, because of these complications, I will avoid the expression “good for” in my thesis.

Hampton sometimes speaks of a reason’s authority as its *normativity* (Hampton 1998, 80, 85). This expression is also used by Christine Korsgaard in her book *The Sources of Normativity* (Korsgaard 1996; see esp. lecture 1). I do not object to this way of speaking. However, in order to avoid misunderstandings, it is important to distinguish between different senses of normativity.

Taken in a wide sense there is normativity wherever there is a standard or rule. Many standards aren’t meant to be applied to actions. This is, as we have seen, true of epistemological standards. Epistemology is not a normative discipline in what we might call the action-guiding sense. Within the category of action-guiding standards, we can distinguish between those that are weakly and strongly normative. Any standard that tells any agent what to do is weakly normative. The official doctrine of the NSDAP meets this description; so does a recipe for tiramisu. These standards are clearly not independent sources of authoritative reasons. They are thus not strongly normative.¹⁷ The Nazis presumably took their ideology to

¹⁶ The agent in question need not even be actual. One could hold that, say, the solar system has intrinsic value even if there is not much any actual agent can do about it. However, it would still have to be true that *if* an agent had the ability to destroy the solar system, he would have reason not to.

¹⁷ Derek Parfit makes a similar distinction between “rule-involving” and “reason-involving” normativity (Parfit forthcoming, 154).

be strongly normative, but they were mistaken. No-one, I take it, has had such grandiose hopes for a tiramisu recipe. In this thesis, I claim that hedonistic egoism is the correct view on strong normativity.

It is often assumed that there is some relation between what one has reason to do and what is *rational* to do. For instance, when egoism is presented as a theory of reasons, it is often called rational egoism (and I will follow this usage in my thesis). However, there is no general agreement on the nature of this relation, and philosophers have also taken the notion of rationality to refer to other things than responsiveness to reasons for action.

First, we talk about rationality in the context of beliefs. Here, being rational amounts, roughly, to forming beliefs that are likely to be true given one's evidence. I will refer to this kind of rationality as epistemic rationality. One can be epistemologically rational or irrational independently of whether one does what one has most reason to do. To illustrate, people who believe in astrology are typically irrational in the epistemic sense even if their belief gives them a lot of pleasure and for that reason would be approved by hedonistic egoism.

Even when it comes to actions, the notion of rationality is used in widely different ways. Sometimes being rational is simply taken to be synonymous with promoting one's self-interest, while in other cases it seems to mean nothing over and above doing what best fulfils one's aims (see S. Rachels 2009, sect. 1, for a discussion of these uses). If rationally is defined in either of these ways, it becomes a substantial question whether there is any reason to be rational. I will simply disregard these uses of the concept.

On yet another view, a person is rational if there is a certain correspondence between his beliefs and actions. Somewhat more precisely, the idea is that an agent is rationally required to intend to do what he thinks he ought to do, and then do what he intends to do.¹⁸ Failing to comply with this requirement makes the agent irrational. It is important to note that no substantive restrictions are being put on the agent's beliefs about what reasons he has; they may be completely mistaken. Thus, Scanlon is prepared to say that if a person sincerely thinks that there is a reason to disregard all future pains that occur on a Tuesday, he is not irrational in actually doing so (Scanlon 1998, 1.4).

Let us call the kind of rationality that amounts to a correspondence between one's normative beliefs and actions *practical* rationality. According to hedonistic egoism, being practically rational is not intrinsically valuable. Imagine that you could choose between two

¹⁸ For a more detailed account of this requirement, see Broome 1999. Broome thinks that acting in accordance with it is a necessary but not sufficient condition for being rational.

lives. In the first life you are supremely practically rational: you always do what you think you ought to do. In the second life you act on irresponsible impulses all the time and suffer from a chronically weak will, but because you live on a very nice planet you nevertheless end up with slightly more pleasure than in the first. Hedonistic egoism would then tell you to prefer the second life. On other views, there may be independent authoritative reasons for being practically rational.¹⁹ In any case, the important point for now is that this is a substantive question about what reasons for action there are.

Finally, there are authors who think being rational in the practical domain includes not only practical rationality, but also being responsive to the correct reasons for action (see e.g. Shafer-Landau 2003, 168). A rational person has correct beliefs about what ought to be done, given a certain set of facts, and then does it. He may of course be mistaken about what the facts are, but that would not prevent him from being rational in the present sense: what matters is that he does what he would have reason to do if his factual beliefs were true. Parfit endorses this understanding of rationality when he says that “we act *rationally* if we act in some way because we have beliefs about the relevant facts whose truth would give us sufficient reasons to act in this way and *irrationally* if we act in some way despite having beliefs whose truth would give us clear and strongly decisive reasons *not* to act in this way.” (Parfit forthcoming, 70)

Scanlon spends some time arguing that his conception of irrationality as a divergence between a person’s normative beliefs and action is preferable to the one espoused by Parfit. Russ Shafer-Landau, on the other hand, thinks we should understand rationality in terms of responsiveness to reasons and claims that Scanlon’s account fails to capture “the very idea of rationality” (Shafer-Landau 2003, 168-9).²⁰ This is a senseless disagreement. There is no such thing as “the very idea of rationality”; we have used one word to refer to (among others) two importantly different things. We should not argue about what rationality really is, but simply distinguish.

Given Parfit’s understanding of rationality, hedonistic egoism as defended in this thesis would count as a theory of rationality, or at least as a theory of what it is rational to do. The theory would hold that it is always rational to do what, given that one’s factual beliefs are

¹⁹ For a general argument that there are no such reasons, see Kolodny 2005. For a defense of the opposite claim, see Southwood 2007.

²⁰ In fact – and somewhat curiously – Scanlon’s view on rationality is different from his view on irrationality. He thinks that rationality involves both being responsive to the correct reasons and acting on one’s beliefs about reasons (see Scanlon 1998, 1.5).

correct, would maximize one's future hedonic level. However, there seems to be no positive need for formulating the theory in these terms, and given all the other understandings of rationality just mentioned, doing so is potentially misleading. I will, therefore, avoid terms "rational" and "irrational" when presenting my views and arguments in later chapters.

In this section, I have tried to make the terminological stipulations and clarifications necessary for a meaningful discussion. To sum up: this thesis argues that only egoistic hedonic reasons have independent authority, that one all-things-considered ought to maximize one's own future hedonic level, that pleasure is the only thing that is intrinsically good and pain is the only thing that is intrinsically bad, and that only the hedonistic egoistic standard is strongly normative.

2.3 Reasons and desires

Reasons for action have also been analyzed in terms of *desires*. Here, too, it is possible to make the definition work by stipulation. One can say that there is an authoritative reason to do what an ideal agent would desire, and then include knowledge of the correct authoritative derivative reasons among the characteristics of an ideal agent. Or, if "desirable" is taken to mean "worth having", then presumably what is desirable is simply what has intrinsic value. It is in this sense I will use the term "desirable" in later chapters of this thesis.

However, some philosophers have thought that reasons can be analyzed in terms of desires in a more interesting sense: reasons are somehow reducible to or created by actual or hypothetical desires. In this section, I will review some criticisms of this idea. I will argue that while the objections do put desire-theories in a questionable light, the matter cannot be completely settled without a closer look at a particular class of reasons, namely hedonic reasons. If we assume, as many authors do, that there are hedonic reasons and that hedonic states are generated or partly constituted by desires, we would seem to have an example of a case where desires generate reasons. Moreover, if they do so in this case, it is hard to see why they should not also have the ability to do so in other cases. Thus, my claim in this section will be that there is one and only one way to salvage the thesis that desires generate authoritative reasons for action: showing that hedonic states essentially involve desires. (In chapter 3, I argue that they do not, so the final upshot of my discussion will be that desires do not generate reasons for action.)

I will understand desires in a broad sense, so that they include motives, preferences and all sorts of pro-attitudes. The arguments I will present do not depend on details about what

exactly the supposedly reason-generating desire is like, nor about what objects it can have, to whom it provides reasons, or when it does so.

Few would deny that we often have reasons to do what we desire to do. However, we need not have these reasons because of the desires. Perhaps the relation between desires and reasons goes the other way: we (typically) desire to do what we have reason to do independently of whether we desire it or not. For instance, my reasons to pursue a career in philosophy may not have anything to do with the mere fact that I desire to do so, but rather be grounded in facts such as that this career would provide me with enjoyment, knowledge or recognition. The reason I have the desire in the first place may be that I find these things valuable.

Thomas Scanlon and Derek Parfit have both defended the view that reasons are grounded in things that have value, and not our desires concerning these things (Scanlon 1998, 1.9; Parfit forthcoming, ch. 2-4). As they point out, desires can still affect our reasons in more indirect ways. Assume that I initially conclude that being a philosopher and being a medical doctor are equally worthy careers. If I develop a desire for becoming a philosopher, this may enable me to enjoy this career more than I otherwise would have done, and so give me more reason to pursue it. Also, I may find it unpleasant to force myself to do something that I at the moment do not desire to do. Even if the desire has no direct hedonic consequences, it may, especially if it is of the kind commonly referred to as a gut feeling, serve as *evidence* that I would enjoy philosophy more than medicine. In any case, it remains true that the reason for wanting an enjoyable career could be completely desire-independent. Furthermore, having unfulfilled desires is typically distracting, and therefore prevents us from doing what we have most desire-independent reason to do. This gives us another kind of instrumental reason for fulfilling our desires. Finally, acting on one's desires can be a useful decision-procedure in situations where little value is at stake, or when too much deliberation would interfere with what we are doing.

It seems, then, that in those cases when desires appear to give rise to reasons for action, there is typically some candidate for a desire-independent reason lurking in the background. In order to find out whether reasons for action are grounded in desires, we should consider desires for things that cannot plausibly be said to have desire-independent value. Imagine, for instance, a person who has an overriding desire to count the leaves of grass in various geometrically shaped areas such as park squares and well-trimmed lawns. The person is (otherwise) intelligent and perfectly normal, and the desire is not based on any false factual beliefs about what this activity is like or the effects it will bring about. He knows that it will

not make him rich, famous, respected, virtuous or any such thing. Would this person have an authoritative reason to spend as much of his life as possible counting leaves of grass?

Now desire-theorists typically require not only that desires must not be based on false factual suppositions, but also that they must in some sense be reflective or considered. Richard Brandt claims that only desires that would survive cognitive psychotherapy have independent normative force (Brandt 1979, ch. VI). Bernard Williams stresses that one's motivational set can be changed by imaginative deliberation on one's actual and possible desires (B. Williams 1979). Michael Smith thinks that we should also adjust our desires, sometimes quite radically, in the light of coherentist reasoning (M. Smith 1994, 158-61). In fact, he thinks that only if it turns out that all agents would agree after going through this reasoning are we entitled to treat the resulting desires as sources of genuine normative reasons (ibid., 164-74). It might seem as if a desire to count leaves of grass could not possibly survive these kinds of idealization. However, if there are no desire-independent reasons for action, there is in fact no basis for thinking that it would not (see Parfit forthcoming, 104-7). A desire for counting leaves of grass might be extinguished by cognitive psychotherapy or imaginative deliberation or coherentist reasoning if these processes involved realizing that the desire-theory is wrong and grasping the desire-independent fact that there is no intrinsic value in counting leaves of grass. But if the theory is correct, why should this particular desire be singled out as necessarily unreasonable? I can think of no good answer to this question.

The desire-theory commits us, then, to saying that the imagined man would have overriding reason to count as many leaves of grass as possible. Some people are presumably willing to accept this implication of the view, and to hold, more generally, that desires generate reasons independently of whether their fulfillment would bring about things that could be thought to have desire-independent value. I have a suspicion, though, that they are willing to say these things partly because they are tacitly assuming that the fulfillment of a desire itself involves something that may have desire-independent value, namely enjoyment (and also that not getting what you desire is unpleasant).

John Rawls, who came up with the original version of the example I have been discussing, may have been guilty of this mistake. According to Rawls, a good life is simply a life that is lived in accordance with a plan that does not violate certain very formal principles of rational choice and is not based on false factual beliefs (Rawls 1999a, sect. 63-5). He should therefore accept that a person who forms a plan to count the leaves of grass on as many lawns as possible during his life and then goes on to do so would be leading a very good life indeed. Presumably, this means that it is also the kind of life that he has most reason to choose.

Rawls' version of the example, however, does not concern the fulfillment of an overriding desire to count grass. Rather, he talks about a person "whose only *pleasure* is to count leaves of grass" (ibid., 379, emphasis added). It is not particularly implausible to hold that this kind of person should devote as much time as possible to counting grass. But Rawls' definition of the good life commits him to a much stronger claim: even someone who never in the least liked the way it *feels* to count leaves of grass or to have his grass-related desires fulfilled, ought to count leaves of grass as long as he desires the counting itself. It would be interesting to know whether Rawls would have agreed with this claim as well.

A supporter of the desire-theory must be willing to accept that desires ground reasons even when they involve no hedonic benefits. Indeed, there must even be a reason for acting on desires for things that are hedonically bad. Parfit objects to the desire-theory on this count, and I have nothing to add to his remarks:

Suppose that [...] I want to *have* some future period of agony. I am not a masochist, who wants this pain as a means to sexual pleasure. Nor am I repentant sinner, who wants this pain as deserved punishment for my sins. Nor do I have any other present desire or aim that would be fulfilled by my future agony. I want this agony as an end, or for its own sake. I have no other present desire or aim whose fulfilment would be prevented either by this agony, or by my having my desire to have this agony. After ideal deliberation, I decide to cause myself to have this future agony, if I can.

Subjective theories here imply that I have a decisive reason to fulfil my desire and act on my decision, by causing myself to be in agony. If there is a fire nearby, and I shall have no other way to fulfil my desire, I would have a decisive reason to thrust my hand into this fire. That is hard to believe. [...]

Though it is conceivable that someone might want future agony for its own sake, this case *is* hard to imagine. This fact may seem to weaken this objection to subjective theories. The opposite is true. This fact *strengthens* this objection. If we find it hard to imagine that anyone might have this desire, that is because we assume what objective theories claim. The nature of agony, we believe, gives everyone very strong reasons to want *not* to be in this state. According to subjective theories, we have no such object-given reasons. If that were true, it would *not* be hard to imagine that someone might want, for its own sake, to have some future period of agony. We could at most claim that this desire would be unusual, like the bizarre sexual desires that some people have. This case is hard to imagine because the awfulness of agony gives everyone such clear and strong reasons *not* to have this desire. It is hard to believe that anyone could be so irrational. (Parfit forthcoming, 109-10)

Until now, I have discussed the view that *only* desires ground reasons for action. One might, of course, also hold that desires are merely one source of reasons among others. Ruth Chang has defended this more moderate theory. She accepts that there are desire-independent or "value-based" reasons, but she thinks there are also reasons to act on one's desires, so that we sometimes have most reason not pick the most valuable option. "I might" she says, "have

most value-based reason to be a doctor, but the fact that I have an affective desire to be a philosopher may give me, all things considered, reason to be one.” (Chang 2004, 86) Importantly, the affective desire does not derive its normative force (solely) from its hedonic nature or consequences; it is supposed to be an independent source of reason for actions (ibid., 80-1).

Chang’s view may seem capable of avoiding some of the objections discussed above. For instance, if there is a desire-independent reason to avoid pain, then this reason may outweigh the desire-based reason the person in Parfit’s example has to thrust his hand in the fire. However, one may still wonder whether there is any authoritative reason to put the hand in the fire in the first place. Similarly, it is less extravagant to claim that the man with a strong desire to count leaves of grass should indulge in this activity a couple of hours a week than to insist that it should be the central focus of his whole life. But those two hours still seem completely wasted. Or imagine that you have followed Chang’s advice and, just because you feel like it, chosen philosophy over medicine even if a career in medicine would be preferable from a desire-independent point of view. You meet a person who, facing exactly the same choice, decided to go against his desire and study medicine. You say to him: “You may enjoy your work somewhat more than I do, you may earn a somewhat better salary, you may do somewhat more good for others, you may have somewhat more knowledge, and so on, but all that is OK because I did what I just felt like and you didn’t. Ha!” This does not seem like a sensible thing to say.

There are, then, strong reasons to be skeptical of the idea that desires generate reasons. However, note that the most convincing arguments all rely on purportedly desire-independent hedonic reasons. It is by no means obvious that hedonic reasons have this status. In fact, the dominant view among contemporary philosophers is that pleasures should be understood, roughly, as experiences that we desire to have for their own sake, while pains are those experiences that we desire to avoid for their own sake. If this is true, the desire-theorist may not be vulnerable to the objection that he is tacitly assuming that the fulfillment of desire involves desire-independently good pleasure, and he can question whether it makes sense to imagine someone desiring his own pain. On a more positive note, he may ask his opponent why, if desires generate reasons in the case of pleasure and pain (and all parties to the debate would agree that there are hedonic reasons), they should not do so across the board (see Rachels 2000, 187-8).

Parfit is aware of this problem.²¹ The simplest solution would of course be to deny that desires play a constitutive role in hedonic experiences. This is what I will do in the next chapter. More precisely, I will argue that pleasures and pains are characterized by certain intrinsic phenomenal feels that are good (or bad) independently of whether they are desired or not. Parfit rejects this account of hedonic states. He thinks that pleasure occurs when sensations are being liked for their own sake at the time they are experienced. However, he insists that these hedonic likings differ from other desires; they constitute a “separate category”. Thus, the fact that hedonic likes and dislikes have the ability to make conscious states good or bad need not mean that desires in general have this ability (Parfit forthcoming, 85-7).

Parfit’s view is not inconsistent, but it raises an obvious question: why are hedonic likings special? Parfit’s only attempt at addressing this question appears to be an observation that they have a different temporal structure: they must be directed at sensations that are presently being experienced, while other desires typically are directed towards the future. One may doubt whether this temporal difference is sufficiently important to carry the normative weight Parfit wants it to carry (see Chang 2004, 77-9). However, the most serious problem for Parfit is that we like all sorts of occurrent states, and not only our present sensations. I might, for instance, like being kind to my mother, having philosophical knowledge or being a citizen of a well-ordered liberal democracy. The man in the example above might like that he is counting leaves of grass (and not in the sense that he enjoys doing so, but in the sense that he likes the mere fact that he is engaged in the activity). Why should only the liking of sensations have an independent capacity to generate reasons for action? This restriction is unacceptably arbitrary. It is like the view that liking only generates intrinsic value when the object of the liking is yellow.²²

Of course Parfit could try to provide some explanation for why hedonic likings have this special status. One possibility would be that there is something about (some) sensations that *merits* liking and is absent in other objects. However, it is hard to see what that could be other

²¹ So is Scanlon, though it is somewhat unclear what his solution is (see Sobel 2005 for a discussion).

²² Fred Feldman, who analyzes pleasure in terms of a propositional pro-attitude, is perfectly willing to accept that this attitude can be directed towards any kind of object, and not just one’s own mental states. One might, for instance, like the fact that the world is at peace, and this liking would (provided it is equally intense) bring about just as much value as a liking of the taste of chocolate or the feeling of being loved (Feldman 2004, ch. 4). I criticize Feldman’s understanding of pleasure in section 3.4.

than an intrinsically good feel.²³ Alternatively, Parfit may think that there is something special about hedonic liking itself, compared to liking with other objects. An obvious suggestion would be that hedonic liking *feels good*. But that would, again, rely on a prior, like-independent notion of pleasantness. Apart from these possibilities, how could hedonic liking be different from other kinds of liking? Parfit doesn't tell us.

I conclude, then, that the plausibility of the idea that desires generate reasons crucially depends on whether desires play a constitutive role in hedonic states, and that this means that Scanlon's and Parfit's objections to the desire-theory as they stand are flawed, or at least inconclusive. However, if we accept the view on pleasure and pain that I defend in the next chapter, the objections I have surveyed do seem convincing, and we should abandon the view that desires are independent sources of authoritative reasons.

2.4 Ultimate reasons

I will assume, then, that if there are authoritative reasons for action, they must be independent of our motives and attitudes concerning them.²⁴ Parfit calls such reasons "objective" or "object-given" and those who believe in them "objectivists about reasons" (Parfit forthcoming, 78). This terminology is not ideal for formulating the view defended in this thesis. Hedonic experiences do count as objects in Parfit's sense, but they are also typically referred to as subjective states. Furthermore, calling a reason objective may, in my experience, make some people think that it is supposed to have authority over all agents. Egoistic reasons do not, of course, have this kind of authority. They can only be objective in the sense given by what I earlier called the requirement of universal validity: if an agent has an egoistic reason to do X, it is an objective fact that he has a reason to do X. For these reasons, I will avoid the terms "objective" and "subjective". I will refer to reasons for action that are non-derivative, authoritative and desire-independent as *ultimate* reasons.

In discussions of normative realism, it is not always clear whether authors are talking about the requirement of universal validity or the stronger requirement that what one ought to

²³ In fact, Parfit explicitly denies that there are phenomenal feels that deserve to be hedonically liked (op. cit., 85).

²⁴ Egoism is sometimes taken to be incompatible with this outlook on reasons. John Kekes, for instance, claims that egoists believe that when someone sincerely thinks he is happy, then he really is happy (Kekes 1982, 368-9). This is simply a mistake. Jan Österberg also ascribes a subjectivist, preferentialist viewpoint to egoism, but this is meant as a claim about the historically speaking dominant ethos behind the theory, rather than a definitional point (Österberg 1988, 39, 2.4, 171-4).

do cannot depend directly on one's desires and attitudes. There is no doubt, however, that if both these requirements can be fulfilled, then realism is vindicated. To say that it is an objective fact that all agents have reason to do what maximizes their future hedonic level, independently of whether they are motivated to do so, prefer doing so, believe they ought to do so etc., is to make a realist claim on any interesting understanding of realism.

The claim that there are ultimate reasons is, of course, controversial. Perhaps some philosophers have supported the desire-theory because they think it is false. This is not a good reason to support the theory, however. It is not given a priori that there are any authoritative reasons for action. If there are no desire-independent reasons, we could accept nihilism about authoritative reasons instead of the desire-theory. If the arguments discussed in the previous section are sound, then this is in fact what we would have (epistemic) reason to do.

There are two kinds of skepticism about ultimate reasons. Some authors seem to think that the search for ultimate reasons is misguided or doomed to failure from the outset: there simply could not be any such things. Bernard Williams' critique of the notion of an external reason belongs in this category (B. Williams 1979). These kinds of criticism have, in my opinion, been handled convincingly by a number of authors (see e.g. Hampton 1998, ch. 2; Crisp 2006, ch. 2; Parfit forthcoming, sect. 15), and I see no need to repeat their points here.

There is, however, a more reasonable and more well-founded form of skepticism towards ultimate reasons. One might accept that it is a good idea to try to develop theories of ultimate reasons, and that it would be important if such a theory turned out to be successful, but doubt that this is in fact going to happen. One may, for instance, be dismayed by the persistent disagreement on normative questions, the metaphysical and epistemological problems raised by claims about ultimate reasons, or the possibility of debunking explanations of our normative convictions. The only way to refute this kind of skepticism is, of course, to try to develop a theory that is able to handle all these problems. Doing so, or at least taking some first steps towards this goal, is the aim of the following chapters.

3. Pleasure and Pain

In this chapter, I discuss the nature of pleasure and pain. I will defend the claim that hedonic states are characterized by their phenomenological feels, and not by being the objects of certain preferences, likes or desires. This claim play will play a crucial role in my argument for hedonistic egoism. As explained in the previous chapter, it is also at the heart of the debate on whether reasons are generated by desires.

3.1 The debate on pleasure and pain

Virtually all the accounts of pleasure and pain in the literature fall into one of two categories. First, some authors think that hedonic experiences contain a certain intrinsic quality – often called a hedonic tone – that accounts for its hedonic status (see G. E. Moore 1903, 64; Broad 1930, 229-33; Schlick 1939, ch. 2; Duncker 1941; Sprigge 1988, ch. V; Tännsjö 1998, 64-7; Crisp 2006, 108-9; Mendola 2006, 443-4; Tännsjö 2007, 84-6). Pleasures all share a positive hedonic tone; pains are experiences with a negative hedonic tone. The standard objection to this view, thought to be decisive by most contemporary writers on the subject, is that there simply is no such thing as hedonic tone – that is, no internal experiential feature common to all pleasures and to all pains.

This has lead many philosophers to believe that the hedonic status of an experience is a function not of its intrinsic “feel”, but of how we respond to it. Pleasures are defined as those mental states that are in some sense desired for their own sake, while pains are those mental states that are correspondingly disliked or avoided (see Alston 1967; Brandt 1979, ch. II; Korsgaard 1996, 145-55; Sumner 1996, ch. 4; Feldman 2004, ch. 4; Heathwood 2007; Swenson 2009, 2.2).²⁵ Apart from this, they need not have anything in common.

In this chapter, I argue that this move to desire-based accounts of pleasure and pain is unwarranted. First, as I explain in sections 3.2 and 3.3, the hedonic tone view has been dismissed too quickly: there is no clear introspective or neuroscientific evidence that pleasures do not share a phenomenal quality. Second, even if it should turn out that they do not, that does not imply that pleasure must be analyzed in terms of any kind of desire. In section 3.4, I try to show that doing so yields a view that is phenomenologically implausible.

²⁵ Henry Sidgwick is often considered the originator of this view (see e.g. Feldman 1997, 448-50; A. Moore 2010, sect. 2), but it is somewhat unclear whether he actually held it (see Sidgwick 1907, 127-31; Crisp 2007, sect. 3).

A more reasonable response is to say that there are several *kinds* of hedonic tone, i.e. several intrinsic “feels” that can make an experience pleasant or painful. This pluralistic hedonic tone view has generally been overlooked in the literature, but appears to be a perfectly possible and even reasonably plausible account of pleasure and pain. I describe it more carefully in the final section of the chapter.

Before I turn to these matters, however, I need to say a few words to clarify the question which the theories just mentioned are trying to answer. In everyday language, the words “pleasure” and “pain” are paradigmatically used to describe certain sensory experiences, such as orgasms and headaches. In the philosophical discussion of which this chapter is a part, however, they are typically used in a broader sense, in which they refer to all experiences that are (in some sense to be discussed in the following sections) positive or negative in their own right.²⁶ For this reason, some writers have thought that we should instead rely on the terms “enjoyment” and “suffering” (see Sumner 1996, ch. 4; Crisp 2006, 101-2). This suggestion clearly has some merit, and I will often follow it. However, to avoid repetition, I will also use the traditional “pleasure” and “pain”. For the sake of simplicity I will sometimes formulate my arguments only in terms of pleasure or only in terms of pain. I do, however, intend them to apply to both negative and positive hedonic states.

Somewhat confusingly, there is also another phenomenon discussed by philosophers under the heading of “pain” (sometimes with the prefixes “sensory” or “physical” added). Human beings and most other animals are equipped with specific nerve-endings, so-called nociceptors, which respond to stimuli that tend to cause tissue damage. This is simply a separate sensory capacity, and – just as we find with all the other sensory modalities (such as sight, hearing, touch, heat etc.) – it involves a particular sensory phenomenal quality. There is something common to experiences of pinpricks, toothaches and muscle cramps that is lacking in other unpleasant sensory experiences such as nausea or smelling rotten meat. Let us call the sensory quality of nociception pain_n and pain in the sense of suffering pain_s (a similar distinction is drawn in Aydede 2000, 546-52; Grahek 2001; Clark 2005, sect. IV-V). Pain_n is in most cases accompanied by pain_s ; otherwise we would of course not refer to them by the same word. The correlation is not perfect, however. Mild pain_n , such as you may experience during a massage, often does not lead to pain_s ; it may even be pleasant. Moreover, there are

²⁶ For instance, Sidgwick makes it clear that he intends his concept of pleasure to apply to “every species of “delight”, “enjoyment” or “satisfaction” [...], the most refined and subtle intellectual and emotional gratifications, no less than the coarser and more definite sensual enjoyments” (Sidgwick 1907, 93, 127).

people who, due to brain damage, appear capable of experiencing intense pain_n without any pain_s at all (see Grahek 2001, ch. 7). Pain_n is, in other words, not essentially hedonically negative. That is why it makes sense to say that sensations of pain need not be painful, or that someone might enjoy feeling pain. In the rest of the thesis, the term “pain” refers (unless otherwise stated) only to pain_s.²⁷

3.2 The traditional hedonic tone view

According to the traditional hedonic tone view, pleasant experiences are characterized by a particular phenomenal quality referred to as positive hedonic tone, while the defining characteristic of painful experiences is considered to be, as one might expect, their negative hedonic tone. It is important not to confuse the two hedonic tones with any particular bodily sensation, perception, thought, volition or emotion. Supporters of the view accept that as long as we are talking about *these* kinds of mental states, we will indeed fail to find one that is shared by all pleasures and all pains. Rather, the idea is that all pleasant experiences – whatever sensations, perception, thoughts, volitions and emotions they are composed of – contain a further ingredient that makes them feel good: the positive hedonic tone. Hedonic tone is, in other words, conceived of as “something ‘over and above’ (other) emotive, conative and cognitive parts of the experience” (Tännsjö 1998; see also Broad 1930, 229-31; Crisp 2006, 108-9; Tännsjö 2007, 84-6). This further “something” has been described as a kind of “glow” or “aura” that “pervades” the experience (Duncker 1941, 399-400). Of course we cannot take this too literally. Supporters of the hedonic tone view admit that it is impossible to give a reductive or non-metaphorical analysis of the phenomenon of hedonic tone; it is conceived of as, in C.D. Broad’s words, a quality “which we cannot define but are perfectly acquainted with.” (Broad 1930)

According to one version of the hedonic tone view, defended by Karl Duncker, hedonic tone only exists as an “aspect” of otherwise non-hedonic elements of consciousness. It is never an independent “unit of experience” in the way that a thought or a sensation is. For instance, when a person enjoys being tickled, the pleasure will not be a separate mental event, but only an aspect or tone of the tickling sensation (Duncker 1941).²⁸

²⁷ The so-called representational theory of pain, according to which pain is just a way of representing actual or potential bodily damage, is probably best understood as an account of pain_n rather than pain_s (see e.g. Tye 2005, 112). If it is intended as an analysis of pain_s, it seems very implausible (see S. Rachels 2000, sect. 2). There is more to suffering than the representation of some fact about the organism.

²⁸ T. L. S. Sprigge also seems to support a version of this view (see Sprigge 1988, 146-8).

This view has some plausibility when it comes to simple sensory pleasures. In these cases it is very difficult to distinguish phenomenologically between the sensory and hedonic elements of the experience, and it may seem as if the latter could not exist apart from the former. However, there are cases where there seems to be no correlation between our hedonic level and any particular thought, sensation or emotion: we simply find ourselves feeling generally good or bad for no obvious reason. Until we are given an explanation of what this kind of apparently free-floating enjoyment or suffering is a mere aspect of, we should probably be skeptical of Duncker's version of the hedonic tone view. Also, the very notion of a mental quality being a mere aspect that cannot exist on its own seems to require some elaboration.

Perhaps it is more promising, then, to suppose that pleasure is some sort of independent "unit of experience", or a mental kind in its own right, rather than just an aspect of other mental states.²⁹ This also seems to be the picture suggested by various findings in clinical psychology and neuroscience (see Aydede 2000; Grahek, 2001, ch. 7; Berridge 2003; Russell 2003). If hedonic tone is a separate and distinctive phenomenal category in this sense, it might actually be a bit misleading to refer to it as a tone in the first place. Fred Feldman thinks we should call it a "distinctive feeling" and reserve the label "hedonic tone view" for accounts of Duncker's type (Feldman 2001, 663). This is a sensible terminological proposal. For the sake of simplicity, however, I will take "hedonic tone view" to cover both these understandings of pleasure. (This is also what the proponents of the view seem to do.) In the context of this thesis, what they have in common is much more important than what sets them apart. They both claim that there is a specific phenomenal quality or feel that is found in all pleasant and in all painful experiences, and it is this claim I will be evaluating in the next section.

A given mental event need not always be accompanied by the same hedonic tone. For instance, I sometimes enjoy eating olives and sometimes do not. Some people seem to never enjoy it, while others enjoy it much more than I do. The explanation of this phenomenon does not have to be that there are different gustatory sensations involved. It could simply be that one and the same sensation results in different hedonic tones across persons or across time. In particular, it should be noted that nociception, or pain_n, as I called it above, is not always accompanied by negative hedonic tone.

²⁹ Supporters of the pluralist version of the hedonic tone view, which I discuss in section 4, could actually adopt a compromise between the two understandings of hedonic tone: some hedonic tones are mere aspects of non-hedonic states while others have a more independent existence.

The generation of hedonic tone should be understood as an involuntary, non-cognitive and purely causal process. Hedonic tone must not be confused with any kind of explicit or implicit conceptual evaluation of the other elements of the experience (see Aydede 2001, 557-8). It is merely a feel. When I enjoy eating olives, I am not committed to any claim whatsoever about the qualities of the gustatory sensations involved. All I am doing is feeling good. Also, we should not, as we sometimes do, say that the taste itself becomes pleasant. Strictly speaking, the only thing that is pleasant is positive hedonic tone.

We often say that pleasures have objects: I may enjoy or take pleasure in, say, a beautiful view. A supporter of the hedonic tone view could take these expressions at face value and hold that hedonic tones are intentional mental states. However, he may also think that the expressions are somewhat misleading. The visual sensations and the positive hedonic tone occur simultaneously in consciousness, and we experience them as intertwined on a phenomenological level, but the sensations still need not be the *object* of the hedonic tone. Perhaps they just cause it. Of course we may also be thinking or making judgments where the visual sensations figure as an object. These judgments and thoughts may even be the cause of the positive hedonic tone in the first place. Furthermore, the hedonic tone can influence our cognitive processes in various ways, for instance by making us think “this is pleasant”. But considered in itself the hedonic tone could just be *there*, and not about any other part of the experience. Even if we assume that Duncker’s view is right and that the positive hedonic tone is merely an aspect of the thought, this aspect itself presumably does not have to be intentional. I leave these questions about intentionality open; they are too complicated to discuss here and the general plausibility of the hedonic tone view does not seem to depend on precisely how they are answered.

Supporters of the hedonic tone view acknowledge that we often make conceptual evaluations of our experiences. But this is not what makes the experiences pleasant or unpleasant in the first place. It is the other way around: we (typically) evaluate them positively because of their pleasantness and negatively because of their unpleasantness. Also, we are normally strongly motivated to prolong pleasant experiences and to escape from painful ones. But according to the hedonic tone view, this motivation is an effect of the hedonic status of the experience, and not constitutive of it. If, because of some psychological malfunctioning (severe depression, for instance), an experience with a negative hedonic tone does not generate any motive to escape it, it can still be a perfectly genuine pain.

The notion of “liking”, which often makes its appearance in discussions of pleasure, is dangerously ambiguous. Sometimes it is just another word for enjoying and may refer to the

kind of non-conceptual, objectless feeling good hedonic tone is supposed to be. For instance, the neuroscientist Kent Berridge has argued, based on an impressive body of evidence, that pleasure should be understood as a kind of “liking”, rather than “wanting”. This liking is, by itself, “simply a triggered affective state—there is no object of desire or incentive target.” (Berridge 2004, 195; see also Katz 2008) On the other hand, as we will see, some philosophers talk about hedonic liking as a kind of conceptual evaluation or propositional pro-attitude.

If one overlooks this ambiguity, the hedonic tone view may seem very implausible.³⁰ For instance, David Sobel finds it hard to believe that someone experiencing positive hedonic tone would be in a pleasant state even if he did not like the phenomenal quality of this state (Sobel 2005, 444-6). It would of course be absurd to claim that an experience could be pleasant even if it was not liked in the sense that it was enjoyed. However, the hedonic tone view does not have this implication. What it says is that an experience could be pleasant even if there is no positive conceptual judgment or pro-attitude directed towards it. And while this may of course turn out to be mistaken, it is at least not *obviously* mistaken.

3.3 The heterogeneity objection

Many authors have rejected the traditional hedonic tone view. The reason they all cite is the lack of a particular experiential quality common to all pleasures and to all pains (Sidgwick 1907, 127; Alston 1967, 344; Parfit 1984, 493; Griffin 1986, 8; Sprigge 1988, ch. V; Feldman 1997, 461-4; Sobel 2005; Heathwood 2007, 26). Is this so-called heterogeneity objection correct? I am not sure. Compare the pleasure of eating chocolate with the enjoyment of a rigorous and fruitful philosophical discussion, and with the pleasure of feeling loved. Clearly the non-hedonic elements of these experiences are irreducibly different. There is no particular thought, sensation or emotion common to all of them. But could it nevertheless be true that they include the same positive hedonic tone? Is the pleasure itself constant across the various pleasant conscious states? Could perhaps the firing of the very same neurons be responsible for the enjoyable character of all these experiences?

I find this a very difficult question. My own attempts to answer it through introspection have so far been inconclusive. To illustrate, consider two activities that I sometimes find pleasant and sometimes hedonically neutral: listening to Beethoven and playing football. It

³⁰ As we have seen, doing so may also make preferentialism about value or reasons seem more plausible than it is.

does seem to me as if there is some elusive further ingredient or “aura” added in the cases where I enjoy these things. I am not, however, able to say that it is the same ingredient in both cases. But this is not sufficient to discredit the hedonic tone view. Positive hedonic tone does not have to be an easily isolable and identifiable part of our experiences. Perhaps it is, on a phenomenological level, fused together with the non-hedonic elements of experience to such an extent that it is difficult to recognize across pleasant experiences.

A stronger claim, which *would* discredit the hedonic tone view, is that we do have clear introspective overview over what pleasant experiences consist in, and it turns out that the common hedonic tone is not there. I find, however, that I lack this kind of overview.³¹

In most of the literature on this topic we find little help beyond mere assertion: some philosophers claim to know through direct self-observation that their pleasant experiences all have the same hedonic tone and others that they don't (see e.g. Sidgwick 1907, 127; Parfit 1984, 493; Mendola 2006, 443-4). The only proper argument for the heterogeneity of pleasure that I am aware of is epistemic. It runs like this:

- (1) When experiencing pleasure while being engaged in two simultaneous activities, we know immediately which activity is (more) pleasant.
- (2) However, if all pleasant experiences had the same hedonic tone, we could not know this immediately.
- (3) Therefore, the traditional hedonic tone view is false.

As Fred Feldman puts it, if you are eating peanuts and drinking beer, you do not have to perform any experiments in order to decide which taste is more pleasant: you know that immediately. But if the two tastes generate exactly the same hedonic tone, it would seem impossible to know through introspection alone where the hedonic tone you are experiencing comes from. Hence there must be something wrong with the thesis that all pleasure is homogenous (Feldman 1988, 63-4).³²

The problem with this argument is its first premise. It is far from obvious that we have non-inferential knowledge of the relative pleasantness of different activities or sensations. A large number of psychological studies have shown that human beings tend to make mistakes

³¹ For a similar claim, see Smuts forthcoming, sect. 5

³² Feldman's argument is only meant to apply to what he calls the distinctive feeling view. On Duncker's view, we could presumably compare the pleasantness of the two activities directly, since the hedonic tone generated by them would be aspects of different sensations and therefore introspectively distinguishable.

when explaining the basis of their preferences and emotions, and that our self-knowledge is often much more indirect than the phenomenology suggests. It should not come as a surprise if something similar occurs in the case of hedonic attributions; in fact, there is by now strong evidence that it does (see Russell 2003). Perhaps we identify the pleasantness of an activity not by direct introspection, but by noticing how our hedonic level co-varies with the stimuli we are exposed to. In Feldman's case, there would presumably be a fast increase in the hedonic level when you have a sip of beer, followed by a gradual decrease. The same would happen when you eat some peanuts. You would know which taste is more pleasant by noticing which bump in the hedonic level is higher. (Of course this process would be automatic; I do not mean to suggest that we go through explicit reasoning in these cases.) If we assume that there are no changes in the hedonic level at all – that every instant of the beer-drinking and peanut-eating experience is equally pleasant – it becomes much more doubtful whether you would be able to make immediate judgments about which taste you are enjoying more.

Another example will illustrate this. Let's say that a mountain climber has just reached his summit and starts enjoying the physical relaxation, the feeling of accomplishment and the beautiful view. His hedonic level is at a constant high level for the next five minutes. It seems to me that in this case – and others like it – it may very well be impossible to quantify, *through direct introspection*, the contributions from these various elements of the experience to the total enjoyment experienced. On the other hand, if there really were three different pleasures involved, then doing so should be fairly straightforward. So perhaps epistemological considerations count in favour of the traditional hedonic tone view, rather than against it.³³ At the very least it would be premature to conclude that we know, on the basis of introspection alone, that there is no hedonic tone common to all pleasure.

It might be tempting to look to neuroscience for help at this point. Of course it is by no means clear what philosophical conclusions would follow from there being just one hedonic mechanism (or several different hedonic mechanisms) in the brain. However, I will not go into that question now, since the facts are not in yet. We do not currently have anything resembling a complete understanding of the neural basis for pleasure and pain (for discussions, see Berridge 2003, sect. 5; Kringelbach 2005; Leknes & Tracey 2008; Katz 2010, 3.3).

³³ Or, more precisely, in favour of the distinctive feeling view. See note XXX.

In sum, it looks to me as if the traditional hedonic tone view could actually be correct after all. I would not want to insist that it is, though. A number of excellent philosophers have found it phenomenologically unacceptable, so perhaps my introspective observer skills just aren't good enough. At any rate, it is worth asking what the alternative would be.

3.4 Liking and motivation

Almost all of the writers who have rejected the hedonic tone view have thought that pleasures should instead be characterized as experiences which we like, prefer or want for their own sake.³⁴ Sometimes these views are characterized as externalism, with the hedonic tone view being correspondingly labelled internalism. Similarly, some authors say that we must decide whether pleasantness is extrinsic or intrinsic to the experience (see Sumner 1996, 87-91; S. Rachels 2000; Crisp 2006, 4.3; A. Moore 2010). I think these terms may be somewhat misleading. First, as we have seen, hedonic tone is not supposed to be internal to the sensations, thoughts or feelings that are part of a conscious state, but only internal to the pleasant experience as a whole. Second, even if you do not think there is such a thing as hedonic tone, you do not have to say that the favourable response that generates pleasure (whatever it is) is external to the pleasant experience. Fred Feldman, for instance, argues that pleasure should be understood as a propositional pro-attitude (Feldman 2004, ch. 4). This attitude simply *is* pleasure. It is intrinsic to pleasant experiences in exactly the same way as hedonic tone is supposed to be.

I think it is better to say that the crucial issue at stake in the debate is what role desires, broadly understood, should be granted in the philosophical analysis of pleasure. Are they only common effects of pleasure, as supporters of the hedonic tone view would say, or are they essential ingredients or preconditions? In this section, I will try to show that we should avoid the latter idea. For the sake of simplicity I will (mostly) formulate my objections in terms of liking, but I intend them to apply to all kinds of desire-based accounts of hedonic states.

³⁴ Richard B. Brandt and Christine Korsgaard analyze hedonic states in terms of motivation (Brandt 1979, ch. II; Korsgaard 1996, 145-55); Fred Feldman and L. W. Sumner in terms of liking (Feldman 2004, ch. 4; Sumner 1996, ch. 4). Many authors do not seem to distinguish clearly between these two options (see e.g. Alston 1967; Heathwood 2007). Adam Swenson claims that the negative reaction that generates pain contains both an evaluative and a motivational component (Swenson 2009, 2.2). It is not necessary for my purposes to discuss the differences between the liking and motivational accounts in any detail. For some criticisms specifically targeting each of them, see S. Rachels 2000, sect. 3-4.

The claim to consider, then, is that the pleasantness of an experience consists in its being liked for its own sake (at the time it is experienced, of course). As I have said, hedonic tone can be described as a particularly primitive form of liking. If we want to discuss an alternative to the hedonic tone view, we must be careful not to confuse liking with hedonic tone.³⁵ When we say that liking the taste of strawberries makes eating them pleasant, for instance, we must not be imagining that the taste causes a non-conceptual feel of pleasure with which it is phenomenologically intertwined. We must be thinking of something that is somehow more cognitive in nature.

Fred Feldman's propositional attitude is clearly such an alternative to the notion of hedonic tone. The attitude is supposed to be on a par with other propositional attitudes such as belief and hope. It is not characterized by any intrinsic "feel", but only consists in an evaluation of some state of affairs (Feldman 2004, 55-6). There may be other and more plausible understandings of hedonic liking, but the important point for now is to remember that it must be distinct from the kind of non-conceptual enjoying that hedonic tone is supposed to be.

Also, of course, we must not imagine that the liking is *based on* hedonic tone. Hedonic preferences cannot simply be reports or effects of a pleasantness that is already there, as the supporters of the hedonic tone view would say they are. That would not give us an independent account of pleasure.³⁶ Rather, the preferences must be responsible for generating the pleasure in the first place: until there is liking, there is no pleasantness at all.

It seems to me that this approach is not going to work. There is more to genuine enjoyment than having a favourable attitude towards your experience. In order to see this, consider the following example. When I stare at a white wall, I usually experience a hedonically neutral state. Let's say that a person has exactly the same sensory experience when staring at a white wall. Because his brain is wired in an unusual way, however, this experience generates a positive propositional attitude. The attitude itself does not have any intrinsically good feel. Nor does it change the sensations of whiteness or provide them with a

³⁵ Guy Kahane defends an account of pain according to which hedonic dislike is hedonically bad "because of how it intrinsically feels." (Kahane 2009) He never explains exactly what sort of thing hedonic dislike is supposed to be, but what he has in mind seems to be very similar to what I referred to as negative hedonic tone. Elinor Mason suggests that "the attitude of 'taking pleasure in' does itself have a feel to it, and the feel is a primitive – the feel is pleasure." (Mason 2007, 382) Her view seems to be a version of the hedonic tone view according to which hedonic tone always occurs together with or as an aspect of a propositional attitude.

³⁶ This point is also made in Crisp 2006, 107.

positive hedonic tone (there is no such thing). The person is experiencing *exactly* what I am experiencing when staring at a white wall, with the addition of a propositional pro-attitude that has no intrinsically good feel. If Feldman's analysis is correct, this qualifies as a perfectly genuine pleasure.

Imagine another person who experiences exactly the same sensory qualities when staring at the white wall. Because his brain is wired in an unusual way, this experience generates a negative propositional attitude. The attitude itself does not have any intrinsically bad feel. Nor does it change the sensations of whiteness or provide them with a negative hedonic tone (there is no such thing). The person is experiencing *exactly* what I am experiencing when staring at a white wall, with the addition of a negative propositional attitude that does not have any intrinsically bad feel. If Feldman's analysis is correct, this qualifies as a perfectly genuine pain.

Something seems to be missing from this picture. As Andrew Moore puts it, it is "hard to see how merely directing one joyless entity at another might constitute a joyful whole." (Moore 2010, 2.2) The addition of like or dislike, understood as propositional attitude and not as feels, cannot by itself turn an otherwise neutral experience into a pleasure or a pain. All we would end up with is a hedonically neutral state that includes a propositional attitude for which there is no rational basis.³⁷

It might be thought that we should not reject Feldman's theory just because it gives what appears to be the wrong answer in some imaginary and rather bizarre cases. However, the real problem is that, on Feldman's view, *all* pleasure and pain must be like this. There can be no difference in principle between having a positive attitude towards sensations of white and having the same attitude towards the taste of chocolate or feelings of fear. It cannot be said that the latter mental states have some kind of attractive or unattractive experiential quality, or that they merit a certain attitude because of how they feel, or that we would (dis)like them if we were just thinking clearly. Doing so would be relying on a prior notion of (un)pleasantness – presumably something like a hedonic tone. Feldman is committed to holding that any experience is, considered apart from the attitude taken towards them, just as hedonically neutral as staring at a white wall usually is to me. If a propositional attitude is not sufficient to

³⁷ T. L. S. Sprigge makes a similar point when he says that Gilbert Ryle's motivational account of hedonic states "gives a strikingly joyless picture of pleasure" (Sprigge 1988, 131-2). See also Goldstein 1980; S. Rachels 2000, 196; Mason 2007.

make staring at a white wall pleasant, it will not be sufficient to make any other experience pleasant, either.

Perhaps there is a more direct way of grasping the implausibility of the attitudinal view. The next time you experience enjoyment or suffering, see if you can find anything in the experience that resembles a response that is different from a hedonic tone and responsible for turning an otherwise hedonically neutral conscious state into a pleasure or a pain. I suspect that you will fail. The hedonic quality of the experience is just *there*, in the brute and immediate way in which the hedonic tone view says it is. As Stuart Rachels says: “When you twist your ankle or jam your finger, the experience itself seems to hurt; the unpleasantness seems to be right there in it.” (S. Rachels 2000, 196)

I believe that the phenomenological implausibility of the preferential and motivational views constitutes a sufficient ground for rejecting them. I will, however, end this section by mentioning a possible further reason for being skeptical of these views.

Most philosophers believe that – at least under normal circumstances – pleasure is intrinsically good and pain is intrinsically bad. I suspect that all those who have discussed the nature of pleasure would agree with this claim. As argued in section 2.3, it would seem that an analysis of pleasure in terms of liking, combined with the claim that pleasure is a source of reasons, forces us to treat likings of any occurrent states of affairs as independent sources of reasons. Similar considerations apply to the motivational view. For those who find these versions of the desire-theory of reasons plausible, or at least not more problematic than the alternatives, this will of course not count against the claim that pleasure is constituted or generated by liking or motivation. However, many philosophers find these general claims about reasons difficult to accept. For instance, they think it could not be true that an intelligent and otherwise normal person could have most reason to spend all his waking hours counting leaves of grass even if he does not like or desire the way it feels, as long as he really likes or strongly desires the very activity of counting leaves of grass. For these philosophers, the fact that accepting a preferential or motivational analysis of hedonic states commits one to making this type of claims about reasons will, to some extent, count against these analyses.

3.5 The pluralistic hedonic tone view

If the standard alternative to the traditional hedonic tone view is not very attractive, what are we to make of the claim made by Sidgwick and others that there is no experiential feature common to all pleasure and all pain? Are we forced to side with the minority and insist that there must be a common hedonic ingredient, after all? This conclusion would be too hasty. It

is simply a mistake to think that the heterogeneity objection, if valid, shows that experiences cannot have hedonic tone.

The traditional hedonic tone view is actually composed of two distinct claims:

- (1) Experiences are made pleasant or painful by their phenomenal qualities.
- (2) There is just one kind of pleasantness and one kind of painfulness.

These claims are logically independent. There is nothing intrinsic to the idea of a hedonic tone that necessitates hedonic monism. In other words, the natural response to the heterogeneity objection would seem to be to discard (2) and not (1). Instead of rejecting the hedonic tone view as such, one should move on to discuss pluralistic versions of it. However, the authors who have made the heterogeneity objection have – without any explicit discussion or justification – taken their objection to disprove (1) (see Alston 1967, 344-5; Parfit 1984, 493-4; Sumner 1996, 92-4; Feldman 1997, sect. VII; Sobel 2005; Heathwood 2007). Even those who have acknowledged that there are difficulties facing the alternatives have thought that accepting heterogeneity must lead to a rejection of the hedonic tone view. Chris Heathwood, for instance, admits that it is “counterintuitive” to say that a sensory experience is pleasant because it is desired, and not desired because it is pleasant, but he ends up biting the bullet since he thinks the alternative would be the (according to him even more counterintuitive) thesis that all pleasures are homogenous (Heathwood 2007, 38-40; see also Feldman 1997, sect. VII).³⁸

Let us, then, have a closer look at the pluralist version of the hedonic tone view. It is important to distinguish between two forms of pluralism. One, which we may call moderate, was suggested by C. D. Broad and has subsequently been defended by T. L. S. Sprigge and Roger Crisp (Broad 1930, 232-3; Sprigge 1988, ch. V; Crisp 2006, 103-9; see also Kahane 2009, 336). According to this view, positive hedonic tone is not a single homogenous phenomenal feel, but rather a phenomenal kind or genus, with members that are irreducibly different, yet at the same time united by a common experiential core. Pleasure is thus supposed to be analogous to a phenomenological kind such as colour sensations: these sensations are also irreducibly different (compare an experience of red and blue, for instance), while at the same time sharing a common basic phenomenology. As Crisp puts it,

³⁸ Aaron Smuts is not convinced by the heterogeneity objection, but he still accepts the assumption that *if* pleasures are heterogeneous, the hedonic tone view is undermined (Smuts forthcoming).

there is something it is *like* to be experiencing enjoyment, in the same way that there is something it is like to be having an experience of colour. Likewise, there is something it is like to be experiencing a particular kind of enjoyment [...], in the same way that there is something it is like to be having an experience of a particular colour. Enjoyment, then, is best understood using the determinable-determinate distinction, and the mistake in the heterogeneity argument is that it considers only determinates. (Crisp 2006, 109)

For all I know, this could be the correct account of pleasure. Above I tried to show that there are no knock-down arguments against the traditional monistic hedonic tone view, but I am not aware of any knock-down arguments in favour of it, either. However, I doubt that this form of pluralism will satisfy the authors who have insisted on the heterogeneity of pleasure. These authors are not only saying that some aspects of some pleasures are irreducibly different, which is what Crisp and Sprigge are willing to accept. They also make the stronger claim that there is absolutely no phenomenal property that is shared by all pleasant experiences: they really have nothing phenomenal in common (see Sidgwick 1907, 127; Alston 1967, 344; Parfit 1984, 493; Griffin 1986, 8; Sprigge 1988, ch. V; Feldman 1997, 461-4; Sobel 2005; Heathwood 2007, 26).³⁹ And this of course conflicts with the idea that there is “something it is like to be experiencing enjoyment”, i.e. a phenomenal core common to all pleasures.

It seems, then, that we need a stronger form of pluralism in order to please the believers in the heterogeneity of pleasure. We would have to say that there are positive hedonic tones that have no shared phenomenal properties at all; they simply feel completely different. What they have in common – and what explains why we refer to them by the same term – is only their positivity: they are *those phenomenal qualities that are, at least for the person experiencing them, intrinsically and desire-independently good simply in virtue of how they feel*. Negative hedonic tones would be those phenomenal qualities that are, in the same sense, bad.

The idea is thus that there is an irreducibly normative element in the characterization of hedonic states. It is possible that Sidgwick had something along these lines in mind when he defined pleasure as “a feeling which, when experienced by intelligent beings, is at least implicitly apprehended as desirable”. (Sidgwick 1907, 127, see also 129) The normative, strongly pluralistic version of the hedonic tone view is also briefly suggested as a solution to the heterogeneity problem in a paper by Stuart Rachels (S. Rachels 2000, 198).⁴⁰ Apart from

³⁹ It should be noted that some of these authors do not distinguish clearly between the two alternatives, but it seems unlikely from the context that they only want to defend the weaker claim.

⁴⁰ Rachels is discussing unpleasantness, but he presumably intends the suggestion to apply to positive hedonic states as well.

that it has not, as far as I know, been defended anywhere in the (relatively large) literature on pleasure and pain. In fact, with the exception of an objection from Roger Crisp to which I will return to below, it has not even been discussed. In the rest of this section, I will elaborate on the view and try to show that it is worthy of more philosophical attention.

In this thesis, I argue that hedonic states only provide reasons to the person experiencing them, and that the value of a pleasure with a given intensity is independent of what its cause or object is. One need not accept these claims in order to accept the strongly pluralistic hedonic tone view as an account of the nature of pleasure and pain. The normative analysis of hedonic states requires that there is egoistic hedonic value, but it is entirely neutral on the further normative issue of whether there are non-egoistic hedonic reasons. Also, it only needs to assert that pleasures are intrinsically valuable when considered as isolated phenomenal states. It is compatible with the possibility that the picture is more complicated when the phenomenal states occur in conjunction with other states of affairs. For instance, one might hold that it is, on the whole, better to derive positive hedonic tone from listening to Beethoven than from playing football, or that the taking of pleasure in the fact that other people suffer constitutes an organic whole with negative value.

Even if some sort of apprehension of intrinsic value is necessary for an experience to be categorized as pleasant, it need not be necessary for the experience to actually be pleasant. Imagine that you are engaged in solving a philosophical problem, and that this experience is good due to the way it feels. If you do not in any way (beyond experiencing it) notice that it is good due to how it feels, you will not, according to the present view, ever have any reason to describe it as pleasant. But wouldn't it still *be* pleasant? Somewhat surprisingly, Sidgwick seems to have held that pleasures must not only be desirable as feelings, but also be "apprehended as desirable" (Sidgwick 1907, 127). However, a supporter of the strongly pluralistic hedonic tone view need not agree with Sidgwick on this point.⁴¹ I leave the question open, since it is not necessary for my purposes in this chapter to answer it.

It is important to stress that we should not understand the strongly pluralistic hedonic tone view as a straight-forward analysis of our pre-theoretical concepts of pleasure and pain. It would be very implausible to hold that whenever a person correctly describes an experience as pleasant, he is consciously and explicitly declaring it to be good for its own sake. Famously, the question "it is pleasant, but is it good?" has an open feel to it. Rather, the idea would be that we have grouped hedonic states together without a clear idea of what actually

⁴¹ Rachels does not: he calls pleasures that are not attended to "background pleasures" (S. Rachels 2004, sect. 4).

unites them, and that philosophical reflection then reveals the common ingredient to be intrinsic phenomenal (dis)value. Of course, if we have put all the good phenomenal feels together under the heading of pleasure, we must in some sense have apprehended or responded to their goodness. However, we need not have done so at a conscious or conceptual level. As an analogy, consider how people who do not know musical theory still may be able to distinguish between tonal and atonal music.

Furthermore, the normative hedonic tone view need not strike one as, once it is formulated, obviously correct. The reader is unlikely to suddenly exclaim: “Yes, that is what I was doing all the time!” The view will, I suspect, only seem plausible (if at all) after some reflection, and in particular after considering the problems facing the alternatives.

It might seem extravagant to claim that people do not generally know what they are doing when using a word as common as “pleasure”, and that we needn’t even be able to immediately recognize the correct criteria for pleasantness when we are presented with them. However, the persistent disagreement between supporters of the monistic hedonic tone view, the preferentialist view and the motivational view is sufficient to show that this is precisely what we should expect. We really do lack a clear idea of what we are talking about when we talk in hedonic terms; this is why philosophers have found it so hard to agree on the nature of pleasure and pain in the first place.

Thus, the following questions all have, at least initially, an open feel to them:

- It is a pleasant experience, but does it contain a phenomenal feel that is found in all other pleasant experiences?
- It is a pleasant experience, but does it involve a propositional pro-attitude?
- It is a pleasant experience, but does it cause a motive to retain it?

Surely the fact that these questions all seem open does not show that all the standard views on pleasure and pain must be wrong. What counts is how things seem after reflection, not before.

Perhaps the question about the goodness of pleasure will seem in some sense especially open. It might do so simply because the pluralistic normative account is not familiar from the philosophical literature. This is of course not an objection to it. On the other hand, one might think that the question cannot be closed in the way entailed by a normative account of pleasure, since those things that we call pleasures are not in fact good at all. This worry will be addressed in chapter 8, when I defend the claim that hedonic states do matter for their own sake.

One could object to the strongly pluralistic hedonic tone view not on the ground that pleasures are not good due to how they feel, but on the ground that something could be good due to how it feels without being a pleasure. If this is correct, strong pluralism would not provide a complete account of the nature of hedonic states. I will, therefore, spend some time discussing this possibility.

There is a not very interesting reason why one might think that some experiences with intrinsic phenomenal value fail to be pleasant. The paradigm examples of pleasant experiences tend to be prompted by some internal or external stimuli. We enjoy having sex, talking to friends or thinking about philosophical problems. As noted in section 3.2, it may be somewhat misleading to talk about the enjoyment itself as having an object. However, it remains true that we do tend to talk this way. In fact, some dictionaries list “enjoy” as an exclusively transitive verb. This may make some people hesitant to apply the label “enjoyment” to phenomenally good experiences that have no clear object or triggering cause, such as tranquillity or a general feeling of being in a good mood. (This is probably the case to an even greater extent when it comes to the concept of pleasure; it does, for instance, sound a bit odd to say that tranquillity could be pleasant.) If so, a supporter of the strongly pluralist hedonic tone view could just say that enjoyable experiences are those experiences that are intrinsically good due to how they feel and have a salient object or triggering cause.

With this possible qualification, then, a supporter of the strongly pluralist view is committed to holding that any experience that is good due to how it feels is also a pleasure. There are two ways to argue against this claim. First, one can point to actual experiences that are phenomenally good and yet not pleasures. Second, one can insist that a given phenomenal feel could have been intrinsically good and yet not appropriately labelled pleasant. I know (except for the possibility mentioned above) of no convincing example of the first kind. Keep in mind that the experience would have to be good only due to how it feels, and not for some other reason (such as being a display of human perfection), and that it would have to be good in itself, and not just good in virtue of its consequences or its participation in an organic whole.

Roger Crisp is a hedonist and so agrees with me on this point, but he appears to make an objection to Rachels’ suggestion of the second kind. He accepts that it makes sense to propose a strongly pluralistic theory of experiential well-being: two experiences with no phenomenal features in common could still both be good for the person experiencing them due to how they feel. For instance, we could subscribe to a view according to which “well-being consists only in feeling warm and hearing the sound of a buzzing bee, i.e. that these are the only

experiences good for people because of how they feel.” (Crisp 2006, 108-9) However, he thinks that these experiences could still fail to be *hedonically* good, or enjoyable. This leaves an appeal to a particular enjoyable feel as the only way of drawing the line between enjoyable experiences and other experiences that are good because of the way they feel. As he says, if we amend the theory “to claim that what is good for people about these experiences is that they *feel good*, then we appear to be back with a monistic form of internalism about enjoyment.” (ibid.)⁴²

To be sure, our pre-theoretical understanding of the concept of enjoyment does not exclude the possibility that the experience of hearing a buzzing bee could be good due to how it feels without being enjoyable. (Crisp blurs the distinctions somewhat by treating “enjoyment” as synonymous with “feeling good”: it is unclear what is meant by saying that an experience could be intrinsically good due to how it feels without being good due to feeling good.) However, as I have said, this pre-theoretical understanding is radically incomplete. In order to refute a view on the nature of enjoyment, it is not enough to point out that something the view entails to be impossible is not obviously so. One needs to show that what the view entails to be impossible seems perfectly possible even on reflection.

How would we know that hearing the sound of a buzzing bee could be good due to how it feels and yet fail to be recognizable as an episode of enjoyment? One possibility is, of course, that we are convinced that there is a phenomenological feel common to all enjoyment, and that the experience of hearing a buzzing bee would lack this feel. Perhaps all Crisp wants to do in the quoted passage is to assert his conviction that there is a phenomenal core common to all enjoyment. However, the interesting question, and the one I am discussing in this section, is whether a strongly pluralist account of hedonic states is a plausible alternative *given that these states are completely phenomenally heterogeneous*. In order to make the objection stick in this context as well, it would have to be the case that even if some of our actual enjoyable experiences are entirely phenomenally different, we have some positive reason to think that if hearing a buzzing bee (or some other experience) had been good due to how it feels, we would lack any propensity to describe it as pleasant.

I am not sure what this reason would be. Admittedly, it is not exactly easy to imagine that the auditory quality of hearing a buzzing bee could, without any changes or additions, be inherently enjoyable. But I do not find it any less difficult to imagine that it could be intrinsically good in virtue of how it feels. We have, I submit, no understanding of how a

⁴² What Crisp calls a monistic view is what I referred to above as moderate pluralism.

normative claim like that could have been correct. It is like the idea that woollen sweaters could have been intrinsically valuable in virtue of being sweaters made of wool.

Consequently, we cannot claim to know all that much about the imagined phenomenal goodness of this auditory quality. In particular, it seems that we cannot, given that the heterogeneity objection is correct, know that the goodness would not be recognizable as hedonic.

I believe, then, that the strongly pluralistic version of the hedonic tone view deserves to be considered a candidate for a philosophical theory of pleasure and pain. The view even looks fairly attractive. It retains the phenomenological plausibility of the hedonic tone approach and avoids the difficulties of the alternatives, while at the same time being immune to the heterogeneity objection. Now we probably should not have too much confidence in strong pluralism until we have some fairly well-founded ideas about what the different hedonic tones are. One possible answer is inspired by G. H. von Wright's idea that there are three different kinds of pleasure: passive pleasure, which is primarily associated with sensations, active pleasure, which derives from activities one is "keen on doing", and the pleasure of satisfaction, which is what we "feel by getting that which we desire or need or want—irrespective of whether the desired thing by itself gives us pleasure" (von Wright 1963, 64-5). As examples, consider the experiences of enjoying a warm bath, enjoying playing football and enjoying hearing your girl- or boyfriend say "yes" when you propose. Perhaps evolution has given human beings the capacity to experience three different positive hedonic tones along these lines.⁴³ Or perhaps strong pluralism will have to be developed in an entirely different direction, or with a much larger number of different hedonic tones. It may also turn out that the heterogeneity of pleasure is not as deep and radical as it has seemed to some, and that a monistic or only moderately pluralistic version of the hedonic tone view will do. Further philosophical examination, combined with results from psychology and neuroscience (where pleasure and pain are currently receiving a lot of attention) is bound to shed more light on this question. It is not my aim here to anticipate the outcome of these investigations. What I have wanted to show in this chapter is only that this is where the focus of the debate should be. The difficult and interesting question raised by the heterogeneity objection is not whether the

⁴³ Presumably most actual pleasures would contain elements of more than one positive hedonic tone. This would explain, to some extent at least, why the different tones are extremely difficult to identify through introspection.

hedonic tone view should be abandoned, but whether and to what extent it should be developed in a pluralistic form.

In this chapter, I have defended the view that pleasures and pains are not characterized by necessarily involving some form of desire, but by having a certain phenomenology. This is, of course, a significant claim in its own right. In addition, it casts doubt on the view that desires are independent sources of reasons. Finally, and most importantly in the context of this thesis, it constitutes a basis for my argument for hedonistic egoism.

4. Hedonism

This chapter describes the hedonism that I will argue for in later chapters. I start by clarifying the content of hedonism as a theory of ultimate reasons. I then give an account of what I take the practical implications of this theory to be.

4.1 Hedonism as a theory of ultimate reasons

The form of hedonism discussed and defended in this thesis is intended as a purely normative theory. It is misunderstood if it is taken to make or imply any factual claim about the causal role or motivational power of pleasure and pain. In particular, it must not be confused with the descriptive thesis called psychological hedonism. The strongest version of psychological hedonism holds that all human behavior is directly caused by a desire for pleasure or an aversion to pain. This seems to have been Bentham's view (Bentham 1823, 1). If it were true, it would be of limited interest to discuss whether normative hedonism is correct, since all of us would act in accordance with it in any case. Strong psychological hedonism is, however, undoubtedly false. Sidgwick thought that even a simple sensual desire such as hunger is not a desire for the pleasure accompanying eating, but "a direct impulse to the eating of food" (Sidgwick 1907, 45). This is perhaps debatable. But when more complicated and intellectual pursuits are concerned, it seems obvious that a desire may be directly for a non-hedonic object, and not for some accompanying pleasure. For instance, a philosopher searching for the correct theory of, say, the nature of numbers need not in the least be motivated by the positive hedonic tone he will experience if he succeeds: he may simply want to discover the truth.

A weaker version of psychological hedonism holds that any action either aims at pleasure directly or at an object that has been associated with pleasure in the past. Even though the object of the philosopher's desire may be non-hedonic, the desire would not have developed in the first place unless gaining knowledge previously had been correlated with pleasant experiences. This seems to have been Mill's view (Mill 1871, ch. IV). As we will see in chapter 8, a more sophisticated version of Mill's account may be closer to the truth than one might think. I do not think that it covers every single action that has ever been performed by a human being. However, whether it does is a question for psychology and not normative philosophy. For my purposes in this chapter, it is sufficient to note that human beings have a wide variety of desires or preferences that are not, *at the moment they occur*, hedonic.

Normative hedonism can be formulated in various ways. First, it does not have to be intended as a complete theory of ultimate reasons. Quite often it is presented as a theory of well-being, i.e. a theory of what makes a life good for the person leading it (see e.g. Feldman 2004, ch. 1; Crisp 2006, ch. 4). Understood in this narrower sense, hedonism is only a claim about a certain class of reasons. The theory is compatible with the possibility that there are non-hedonic ultimate reasons that are not based on well-being (such as for instance reasons stemming from deontological duties). On some understandings of the concept of well-being, it is even conceivable that there is no ultimate reason to seek well-being (see Crisp 2006, ch. 3). In this thesis, however, I will (unless otherwise noted) only discuss hedonism as a complete theory of both reasons and value.

Hedonism about ultimate reasons is the view that all valid authoritative and non-derivative reasons for action are grounded in hedonic states. This general doctrine can be specified in a number of directions. First, there is the question of whose pleasure and pain it is that provides an agent with ultimate reasons: is it only the agent's own, or are we normatively compelled to be concerned with the hedonic states of any sentient creature? Is the truth perhaps somewhere in between? In chapters 7-9 I will explain why I think the egoistic interpretation of hedonism is most likely to be correct. The account of the practical implications of hedonism in this chapter will, however, remain neutral on this point.

According to what we might call the classical conception of hedonic reasons, since it was espoused by both Bentham and Sidgwick, all pleasure of a given intensity is equally worth seeking, independently of its object or cause. Moreover, one ought to maximize the relevant sum of hedonic value (be it one's own or everyone's), and one ought to do so in a temporally neutral way. All these claims have been disputed. I will defend them later in my thesis, but for now I will simply assume that they are correct.

The claim that one ought to maximize hedonic value presupposes that quantitative comparisons of hedonic value are, at least in principle, possible. Before I go on to discuss whether they are, it should be pointed out that a form of normative hedonism could still be correct even if it turns out that it is completely impossible to quantify hedonic value, or indeed make any kind of value comparisons between different hedonic experiences. It could still be true that only pleasures are intrinsically good and only pains are intrinsically bad; we would just never know how good or how bad. We could not automatically infer from the fact that hedonism would provide little exactness or practical guidance that something other than pleasure must be intrinsically valuable instead. In fact, my argument for hedonism in the final part of this thesis does not at any point presuppose that hedonic value comparisons can be

made. That said, I do want provide some brief arguments in this section for the claim that they can.

The first thing to note is that even if there are objective facts of the matter concerning the relative value of hedonic experiences, human beings would rarely, if ever, be in an epistemological position to make accurate judgments about them. As Sidgwick points out, the basic problem is that when we compare a pleasure with another pleasant experience,

this latter must generally be a representation, not an actual feeling: for though we can sometimes experience two or perhaps more pleasures at once, we are rarely in such cases able to compare them satisfactorily: for either the causes of the two mutually interfere, so that neither reaches its normal degree of intensity; or, more often, the two blend into one state of pleasant consciousness the elements of which we cannot estimate separately. But if it is therefore inevitable that one term at least in our comparison should be an imagined pleasure, we see that there is a possibility of error in any such comparison; for the imagined feeling may not adequately represent the pleasantness of the corresponding actual feeling. (Sidgwick 1907, 141)

Sidgwick goes on to survey the various errors and biases that arise in this way, and he concludes that there is no way of avoiding them entirely (*ibid.*, book I, ch. III-IV).⁴⁴ Stuart Rachels argues – correctly, it seems to me – that more recent attempts at establishing a basis for a precise hedonic calculus fail for the same reasons (S. Rachels 2004, 250-4). Thus, when thinking about the theoretical possibility of hedonic value comparisons, we should keep in mind that even if there are perfectly objective facts of the matter, actual comparative judgments should not be expected to yield more than a rough approximation to the truth.⁴⁵

When it comes to the value of a single episode of pleasure, the standard idea is that it is determined by the product of its average intensity and its duration (see e.g. Feldman 2004, 25-6).⁴⁶ Of course, we must not forget that the relevant intensity is that of the hedonic tone and not that of any other element of the experience. The fact that an intense bodily sensation gives you pleasure does not mean that the pleasure itself must be particularly intense.

It seems fairly obvious that time and intensity are quantitative notions. However, according to Rawls, the idea that they should be multiplied is ultimately just an arbitrary stipulation; the “quantitative dimensions of pleasure, intensity and duration” are mutually

⁴⁴ There is some empirical evidence that people make additional errors in retrospective evaluations of painful experiences (see Broome 1996).

⁴⁵ If we ever understand the neural mechanisms underlying pleasure, and if the mechanisms are of the right kind, the situation might be different.

⁴⁶ What matters is presumably subjective time and not objective (see Tännsjö 1998, 68-9).

“incomparable” (Rawls 1999a, 488). Now it does make sense to doubt whether you should aim for as much pleasure as possible. Perhaps you should instead try to make the most pleasant experiences as pleasant as possible, or the most painful as little painful as possible. But once you have decided to maximize the sum of pleasure in a temporally neutral way – and it is this kind of hedonism Rawls is discussing (see *ibid.*, 486-7) – I find it hard to see how pleasure could be quantified in any other way than by multiplying intensity with duration. The question of whether a maximizing hedonist is to choose a “brief but intense pleasant experience” over “a less intense but longer pleasant experience”, to which Rawls thinks there is no clear answer (*ibid.*, 488), seems to be like the question “if we want a box that is as voluminous as possible, are we to choose one that is really wide but not very tall over one that is really tall but not very wide?”.

As long as the hedonic tone remains the same, intrapersonal comparison between different hedonic experiences poses no theoretical problem. However, as we saw in the last chapter, it is possible that there are several positive and negative hedonic tones with no shared phenomenal properties. On the moderately pluralistic hedonic tone view, there is a phenomenal core common to all pleasures and pains, but there are also irreducible phenomenal differences. Even on the monistic hedonic tone view, it is not obvious that the positive and the negative hedonic tone can be put on a single, homogenous scale.⁴⁷ We need to ask, then, whether and in what sense it would be possible to compare the value of experiences containing irreducibly different hedonic tones.

One possibility is that this cannot be done at all. Rawls, for instance, thinks that pleasures are “incomparable” along their qualitative dimensions as well as their quantitative (*ibid.*). It is worth pausing to note how extreme this claim is. Let’s say, for instance, that I could have one hour of reasonably interesting philosophical conversation if I agreed to forego all sexual pleasure for the rest of my life. If it literally is impossible to compare the hedonic that is typical of sexual pleasure with the hedonic tone that is typical of philosophical conversation, I have no basis for judging what, from a hedonistic point of view, I ought to do in this case. I might as well toss a coin. This is a view that we should not accept unless we are given very strong reasons to do so. I am not aware of any such reasons.

A more moderate position is the view that there is a limited amount of vagueness in hedonic value comparisons. It may be clear that a lifetime of sexual pleasure is better than one

⁴⁷ The research shows that there is heavy interaction between the pleasure- and pain-generating areas of the brain (see Leknes & Tracey 2008), and this may suggest that there is such a scale, but it hardly proves it.

hour of philosophical conversation, but as we reduce the amount of sexual pleasure, we may reach a zone of indeterminacy where its value is neither greater nor smaller than the intellectual pleasure.⁴⁸ This view has a consequence that might seem somewhat puzzling. Let's say that A and B are experiences with the same hedonic tone and intensity; the only difference is that A lasts somewhat longer. However, they are both located in a zone of indeterminacy with respect to an experience C with another hedonic tone. We would then have to say that even though there is no reason to go one way or the other in a choice between A and C or between B and C, there is a reason to choose A over B. Also, in order to avoid counterexamples of the kind presented above, the zone of indeterminacy would have to be fairly small. This raises a question about whether it is necessary to postulate it in the first place. Perhaps the appearance of a zone of indeterminacy is just a reflection of the epistemological shortcomings mentioned above.

One could also hold that it is always, in principle, possible to make exact comparisons of the value of hedonic experiences. For instance, it could be the case that, keeping intensity and duration constant, the value of one hedonic tone is exactly twice as great as the value of another.⁴⁹ It might seem, however, that we should only expect this kind of exactness if there is some more basic phenomenon underlying the different hedonic tones or their value, and it is unclear what this phenomenon could be.

Finally, it is possible that the value of different hedonic tones could be incommensurable – not in the strong sense of being incomparable, but in the sense found in mathematics. It would then be impossible to express the relative values of different hedonic tones by a rational number. However, this would not stop us from making quantitative comparisons; they just could not be completely exact. The ratio between π and the square root of three cannot be expressed by a rational number, but that doesn't stop us from figuring out that the former number is nearly twice as large as the latter. If we are good mathematicians or have access to the necessary technology, we may give a much closer approximation. This kind of incommensurability between hedonic tones would not, in other words, constitute any important theoretical problem for hedonists.⁵⁰

⁴⁸ For a discussion of this and some similar possibilities, see Hsieh 2010.

⁴⁹ In this spirit, Plato offers a demonstration that the life of the philosopher is exactly 729 times more pleasurable than the life of the dictator (*The Republic*, 587b-588a). It is presumably not meant to be taken seriously.

⁵⁰ One might perhaps suspect that the value of experiences containing different hedonic tones can only be compared ordinally, and not cardinally. However, as long as experiences containing one and the same hedonic

With the exception of the first one, these outlooks are all congenial to the form of normative hedonism I defend in this thesis, and I will not spend more time discussing them.

There are further potential problems concerning *interpersonal* hedonic value comparisons. As I explain in chapter 6, rational egoism does not entail that there is never any point in making such comparisons. However, the theory does not presuppose that they can be made, either. I have, therefore, decided not to address the issue of interpersonal hedonic value comparisons in this thesis.

4.2 The implications of hedonism

If all ultimate reasons are hedonic reasons, how should we live? In popular culture, the hedonist is usually portrayed as a person single-mindedly trying to gratify all his sensory impulses.⁵¹ This is obviously not the lifestyle the philosophers who have been hedonists have wanted to recommend. In fact, I will argue that when it comes to how to act, hedonism does not make particularly dramatic claims. It would tell most people, at least within contemporary Western societies, to lead their lives in roughly the same fashion as they do today. On the other hand, it is, as we will see, more of a revisionist doctrine when it comes to how this behavior is justified.

The first mistake in the popular conception of hedonism is the idea that bodily sensations are the primary sources of pleasure. Positive hedonic tone is, as we have seen, not essentially tied to sensory stimulation; it can be triggered by a wide variety of mental events. This point has of course been made a number of times by a number of hedonists. Bentham, for instance, thought that “the pleasures of sense” constitute only one of 14 different kinds of pleasure (Bentham 1823, ch. V). But perhaps we still tend to underestimate how often positive hedonic tone has little or nothing to do with what is going in the body. I will not follow Bentham and attempt to provide a complete categorization of non-sensory pleasure; instead I will just list some experiences I typically enjoy:

- Understanding how the world works
- Discovering, pondering and solving theoretical problems

tone can be compared cardinally, the possibility of ordinal comparisons between different hedonic tones would entail the possibility of cardinal comparisons.

⁵¹ Apparently, this conception of hedonism can also be found among neuroscientists working to understand pleasure. In a recent paper, we are told that “extreme pleasure-seeking and pain avoidance (hedonism) can have undesirable consequences such as drug addiction and obesity.” (Leknes & Tracey 2008)

- Playing poker
- Being in love
- Helping people I care about
- Sharing opinions and frustrations with someone
- Being admired
- Being relieved
- Watching “Monty Python’s Flying Circus”
- Reading sentimental novels
- Feeling full of energy
- Buying something I like at a reasonable price

These are some experiences that I usually find unpleasant:

- Reading a badly written philosophy paper
- Struggling to formulate precisely what is wrong with an argument
- Realizing that few people will ever share my philosophical views
- Worrying about whether I have enough time to fulfill all my commitments
- Being afraid
- Feeling guilty for hurting someone I care about
- Being rejected
- Being embarrassed
- Lacking the energy to do what I have to do
- Being bored
- Not being able to make a decision
- Not being able to sleep
- Realizing that I have deleted an important document
- Being caught in traffic when I am in a hurry
- Having to suppress strong impulses

Needless to say, these lists will differ from person to person. The important point to keep in mind is that they would all, if complete, contain a very large number of activities. Pretty much everything we do can be accompanied by a positive or negative hedonic tone.

It should also be pointed out that some things are hedonically valuable not primarily because they provide pleasure directly, but rather because they influence subsequent hedonic

reactions. For instance, a hedonist would agree with Rawls that self-respect, understood as the conviction that one is a capable person leading a reasonably successful life, is crucial to human happiness (Rawls 1999a, 386). People with self-respect will of course now and then experience a warm, pleasant feeling when they contemplate their accomplishments. But the real hedonic value of self-respect lies in how it affects our hedonic reaction to positive or negative events in our lives. Having self-respect enables you to bounce back faster from failures and to enjoy success more thoroughly. Being involved in well-functioning and close relationships to other people provides similar, but perhaps even more general, indirect hedonic benefits (see e.g. Layard 2005, ch. 5). There are also indirect hedonic harms, of course; a good example is what happens when parents spoil their children.

It should be clear, then, that hedonism does not tell us to be exclusively concerned with simple sensory pleasures. But would it at least tell us to be more concerned than we are today? This is of course an extremely large and complicated empirical issue, but my suspicion is that the answer, in general, is “yes, but only a little bit more”. Most contemporary Westerners are already quite avid seekers of physical pleasure; the once wide-spread conviction that some such pleasures (and in particular those derived from sexual activities) are intrinsically inferior, or even intrinsically bad, seems to be disappearing. Moreover, there are also problems with sensory pleasures: they tend to have bad long-term effects, and our capacity to experience them is fairly limited. A gourmet dinner can provide quite intense pleasure, but not if you have one two times every day. For these reasons it is unlikely that hedonism would recommend a radical shift in the direction of sex, drugs and rock ‘n’ roll. But there are of course people around who needlessly miss out on some quite valuable sensory pleasures.

The second mistake in the popular conception of hedonism is the idea that the hedonist is someone who is preoccupied with pleasure, always making sure that he is at his hedonic maximum. This is not implicit in the hedonist doctrine, as defended in this thesis. Hedonism is a purely normative theory of ultimate reasons, and not an empirical claim about what sort of psychological make-up will generate the most pleasure. Conceivably, being a fervent anti-hedonist could have this effect. As noted in chapter 2, this would not constitute any theoretical objection: even if there is no ultimate reason to become a hedonist, it could still be rational in the epistemological sense, and the philosophical claim that only pleasure has intrinsic value could still be correct.

In ordinary circumstances, being a fervent anti-hedonist is of course not the best way to attain positive hedonic tone. There are, however, several reasons why it is not a good idea to be too concerned with pleasure, either. First, as Sidgwick notes, if we are constantly monitoring the hedonic level of our experience, we may end up reducing it (Sidgwick 1907, 138-40). It is, for instance, usually not a good idea to ask yourself “is this as pleasant as it should be?” while having sex. In addition, this kind of self-monitoring would, in the long run, simply be exhausting. Explicit hedonic evaluations of activities should, therefore, not be made too frequently – and preferably not while the activity is at its most pleasant.

Second, hedonism should not be used as a universal decision procedure.⁵² Hedonic calculation is time-consuming, requires a lot of information and is usually not particularly pleasant. To a large extent we should instead rely on habits and hedonic rules of thumb in everyday deliberation. For instance, under ordinary circumstances I do not stop to ask myself whether I will maximize pleasure by going to work; I just go. Explicit calculation should be reserved for choices that are very simple, important or unusual, and for revising our existing habits and rules of thumb. This strategy also has the benefit that it often protects us from giving in to temptation and choosing a smaller short-time hedonic gain over a long-term loss. Chances are if I did go through with a hedonic calculation every morning, I would stay home a bit too often.

But how, it may be objected, can we know when to engage in hedonic calculation unless we already have engaged in hedonic calculation in order to determine which kind of deliberation is appropriate? Would we not be forced to be calculating all the time after all? I am not sure exactly how we solve this problem, but, as Samuel Scheffler points out, that is a question for psychology (Scheffler 1992, 37). It is simply a fact that humans have this ability – and we use it in many other contexts than when applying hedonism. To deny this is, in the words of R. M. Hare, “like saying that in a battle a commander cannot at the same time be thinking of the details of tactics, the overall aim of victory, and the principles (economy of force, concentration of force, offensive action, etc.) which he has learnt when learning his trade. Good generals do it.” (Hare 1981, 52)

Third, a hedonist should not always be directly motivated by a desire for pleasure or an aversion to pain. When it comes to sensory pleasures, being hedonically motivated is usually not a problem: a piece of chocolate will taste just as good (or perhaps even better) if eaten

⁵² Utilitarians often make a corresponding point about their criterion of moral rightness (see e.g. Hare 1981, ch. 3).

solely for the sake of pleasure. But many pleasant activities are different. For instance, talking to a friend while constantly striving for the enjoyment this activity normally yields is very unlikely to be successful. We get a much better result if we simply act on a friendly impulse. Something similar is true for many of the other non-hedonic projects that keep human beings occupied in everyday life. These projects constitute extremely valuable sources of pleasure: being absorbed in activities can be very enjoyable, it protects us from boredom and lack of energy, and it has the indirect hedonic benefit of promoting self-respect. Of course there is also a downside: having non-hedonic desires will sometimes lead you to make certain hedonic sacrifices. It seems quite obvious, however, that the benefits generally outweigh the costs. Those people who feel most happy tend to be those who are engaged in many other pursuits than feeling happy. A successful person, as judged by hedonistic criteria, is therefore likely to be a person who develops and maintains non-hedonic motives to pursue a wide variety of goods, including those that are often claimed to be intrinsic values in competing theories of ultimate reasons, such as beauty, knowledge, autonomy and friendship.

Of course hedonism doesn't condone just any non-hedonistic project. The pursuit (or the success) of the project must not only provide pleasure; it must provide at least as much pleasure as the available alternatives. I take it that for all existing people hedonism would recommend weakening or extinguishing some desires, strengthening others, and creating a few new ones. However, the extent to which we can influence our desires and motives is limited, and there are many complex factors to take into account when deciding what system of desires is most attractive. There is a limit to how much time and effort one should spend on perfecting one's motivational set. If the existing set works reasonably well, it may be a good idea not to interfere too much with it. It should also be noted that a well-functioning motivational set provides a useful decision procedure. In many situations it is more efficient to simply do what you feel most motivated to do, rather than trying to rely on hedonic calculation, habit or rules of thumb.

This hedonic desirability of non-hedonic motivation is what Sidgwick referred to as "the fundamental paradox of Hedonism" (Sidgwick 1907, 48). The paradox is not a theoretical problem. As pointed out, hedonism is a theory of ultimate reasons, and cannot be falsified by any empirical facts about the causes of pleasure. However, it might seem to constitute a serious practical problem for believers in hedonism. Once we are convinced that pleasure and pain are the only things that matter for their own sake, will we not lose all non-hedonic motivation, thereby being deprived of very important sources of pleasure?

Clearly, the human mind *could* have worked in such a way that we were only motivated to do what we believed we have an ultimate reason to do. Then any desire would have disappeared as soon as we became convinced that its object lacks intrinsic value. However, our psychological make-up is not that neat and hierarchic. It contains a number of more specific impulses that are formed more or less automatically and lie beyond direct conscious control. Without having made any conscious choice we find ourselves caring about some of the things that happen around us, wanting certain states of affairs to obtain and others not to obtain. This is a very deep feature of our psychology, and no philosophical belief will, by itself, make it disappear. (I discuss some of the mechanisms underlying this phenomenon in chapters 7 and 8.)

Curiosity is a good example of this kind of specific and non-voluntary impulse. The drive to learn and discover new things cannot be reduced to a desire for pleasure or any other more “basic” object. It does not arise as a result of a philosophical conviction that knowledge is intrinsically valuable, nor do those of us who are convinced that knowledge is not intrinsically valuable automatically stop being curious. We never decide to be curious; we simply find ourselves having the impulse. Something similar seems to be true of sympathy, envy, phobias and many other human drives. Of course I am not denying that, if you really want to, you may be able to eradicate (or at least drastically weaken) all these non-hedonic impulses. The point is that you would have to actively choose to do so; it would take a lot of time and effort. And people hardly ever have a good hedonistic reason to make that choice.

Games provide another useful illustration of how motivation can be independent of beliefs about intrinsic values. For instance, I sometimes play football. Being a hedonist, I am, of course, convinced that the outcome of football matches do not really matter for their own sake. The value of the activity is entirely derived from the enjoyment it provides to those who participate or watch. Yet this conviction does not stop me from forming a fairly strong non-hedonic desire that my team should win. I would even, in the heat of the moment, be willing to make a (moderate) net hedonic sacrifice in order to secure a victory. As a result, I get, in the long run, much more enjoyment out of football than I would with a purely hedonistic motivation.

It is important to note that the non-hedonic desire is not a result of any kind of self-deception. It is not as if I say to myself: “From a philosophical point of view, it doesn’t matter at all who wins the game. But if I manage to delude myself into thinking that it actually does, I might get quite a lot of pleasure out of playing. So I will temporarily adopt a false theory of ultimate reasons.” This process would never work. (If it did, I might as well get pleasure out

of watching paint dry.) I do not revise my normative philosophical beliefs at all; if someone were to ask me seriously, I would admit that the victory is, in itself, completely worthless. I just find myself desiring it even so.

Another example of this phenomenon is the common non-hedonic desire for money. Perhaps some people really have deluded themselves into believing that money – usually taken to be the paradigm example of an instrumental good – is actually an intrinsic value. But most of us know perfectly well that money is not valuable *per se*, but only because of the non-monetary goods it can bring about. Yet we still feel an attraction towards the very idea of having or earning money. Within limits, this can actually be a good thing as seen from a hedonistic point of view, since it tends to make us more financially prudent and since making money can be an enjoyable project in its own right.

There are a number of complex psychological and sociological factors determining whether a given aim or activity is able to grab our attention in this automatic way. There are also, of course, large interpersonal differences. Some people never learn to enjoy football, for instance. But it seems clear that, in general, all the major non-hedonic candidates for intrinsic values have this ability. If human beings can develop non-hedonic desires for something as mundane as money or victory in sports without engaging in self-deception, they can certainly do the same when it comes to art, science, freedom or friendship.

Of course I do not want to claim that the hedonist will pursue these latter goods just as persistently as someone acting in accordance with value pluralism would. He is unlikely to sacrifice his life for freedom, for instance (though if he is not an egoist he might do so for the hedonic benefits of freedom). A more interesting question is whether hedonism would support a lower general level of commitment to non-hedonic values and projects than what we find among actual people in our society. I suspect that it might – but again it seems to me that the changes recommended would not be very dramatic. Although many people presumably are, to some extent, motivated to pursue non-hedonic goods by a belief in their intrinsic value, it is probably also true that many underestimate how hedonically valuable such projects are. It is therefore unlikely that a substantial reduction in the commitment to non-hedonic aims would have positive hedonic effects.

From a hedonistic point of view, the fact (if it is one) that believing in hedonism would lead to less commitment to non-hedonic goods is of course not a problem. However, it would presumably also lead us to *enjoy* our non-hedonic projects somewhat less. For instance, a hedonist would not derive quite as much pleasure from developing a successful scientific theory as a person who is convinced that knowledge has intrinsic value. And *this* clearly is a

problem from a hedonistic point of view. One may of course doubt how significant the loss in pleasure would be. But it does seem to be an unavoidable consequence of believing in hedonism. In other words, being a hedonist is, to some extent, self-defeating.

Does this mean that most people have most reason (in the sense relevant for action) for not being hedonists? Now I doubt that it is ever a good idea to try to make oneself *stop* believing in hedonism. This would require some very elaborate self-deception which, if at all psychologically feasible, certainly would not be conducive to good mental health. On the other hand, it may be that some people do better, hedonistically speaking, by never having been hedonists in the first place. For instance, a successful artist may perhaps bring more pleasure to himself and others if he believes in the intrinsic value of his work. But for most people under most circumstances, disbelieving hedonism will not produce any major benefits and is likely to lead to substantial sacrifices of pleasure. I think, therefore, that hedonism does in general recommend believing in hedonism.

By now the third mistake in the popular conception of hedonism should be apparent. Although the hedonist's aim is fairly simple, putting the doctrine successfully into practice is anything but. It is not accomplished by mere mindless indulgence, but requires lots of sophisticated reasoning and psychological understanding. It also requires a well-developed capacity for self-control – not only in order to resist pleasures with bad long-term consequences, but also because of the many non-hedonic impulses that now and then need to be checked. For instance, it can be very difficult to curb a desire for professional success even when you know that doing so will make your life more enjoyable.

I have argued that the practical implications of hedonism are, in general, not very radical, as seen from the point of view of contemporary Western democracies. Of course there are, according to hedonism, many serious problems in these societies, but they tend to be widely acknowledged already. Hedonists are not the only ones who worry about things such as pollution, drugs, inequality and consumerism. There are certain specific areas, however, where adopting hedonism would have more profound effects. One is the question of euthanasia. It seems clear that active euthanasia, if properly institutionally implemented, is a hedonically desirable practice, yet it is currently only allowed in a few countries. Another is the treatment of animals. If pleasure and pain really are the only things that matter for their own sake, it becomes much more difficult to justify our relatively lax attitude towards animal suffering in farming and research.

Perhaps the most significant mistake in the policies of Western countries, as judged by hedonistic standards, is the lack of resources devoted to treatment and prevention of mental illness. Mental disorders typically involve very high levels of suffering, and it is believed that, at any given time, about one sixth of us suffer from (at least) one of them. Mental health problems, including addiction, are responsible for roughly 40% of the total amount of disability in the EU and US – and because they tend to be so very unpleasant they are presumably responsible for an even greater share of the total suffering caused by ill health. On the basis of these facts Richard Layard, one of the pioneers in so-called happiness economics, has claimed that “mental illness is probably the single largest cause of misery in Western societies.” (Layard 2005, 181-3) Yet we do relatively little about it, especially compared with the enormous efforts we make to solve other health problems. According to WHO, rich countries spend on average less than 7% of their health budgets on mental health problems (the numbers are substantially lower for poorer countries) – even if there is no doubt that many more of those afflicted could have been helped if the funding had been better (World Health Organization 2003, 37-38). From a hedonistic point of view, this is clearly unjustifiable.

My aim in this chapter has been to explain what the hedonist way of life would be like for most people in most circumstances. In order to make a complete appraisal of the doctrine, it must also be examined what it says about various unusual circumstances. I will postpone this task until chapter 9, where I defend the doctrine from various counter-examples. It should be noted, however, that hedonism does have quite counterintuitive implications in some of these cases. It has not been my intention to try to conceal this fact.

It should also be noted that while the behavioral implications of hedonism may be reasonably similar to those of value pluralism, the philosophical claims made by these doctrines are quite different. According to hedonism, most of the things we do do not really matter; it just so happens that we will enjoy ourselves more if we go on doing them anyway. All you can reasonably hope to get out of your life is a moderate surplus of pleasure over pain, something you may be likelier to experience as a four-year-old than, say, a successful philosopher.⁵³ This does not mean that life is meaningless, but it is certainly less significant

⁵³ The Greek hedonist Hegesias reportedly thought that even this goal is unrealistic, and that the safest policy for a hedonist is to commit suicide (see Matson 1998). In a remarkable passage, Wallace I. Matson argues that Hegesias is right about this implication of hedonism, since hedonism is built on a “game-theoretic model of life”

than many philosophers have wanted to believe. Richard Taylor, for instance, states that a child “is not happy in the sense that is important to philosophy, that is, in the sense of having achieved fulfilment or having been blessed with the highest personal good [...], the kind of good that normally takes the better part of a lifetime to attain.” (Taylor 2002, 111)⁵⁴

Hedonism denies that there is a “highest personal good” of this kind in our lives. It is more in line with Hume’s remark that from a philosophical point of view life “is to be regarded more as a dull pastime than as a serious occupation.” (Hume 1743, 97) Fortunately, it’s not *always* dull, though.

If hedonism is to be an acceptable theory, these claims about ultimate reasons must be correct, and not just capable of generating everyday implications that are by and large not repulsive. Let’s say that a value pluralist criticizes hedonism for failing to accord intrinsic value to knowledge, autonomy and various other non-hedonic goods. Then it is certainly not a satisfactory reply to point out that under most circumstances, hedonism would have you act roughly *as if* these things have intrinsic value. What the value pluralist insists on is that they really do have intrinsic value, and the hedonist has to explain why they don’t. I will, as I have said, attempt to give such explanation in chapters 7-9.

and since game-theory dictates that we should apply the so-called maximin principle when choosing under uncertainty (ibid., 556-7).

⁵⁴ Taylor thinks that the happy life consists of creative use of intelligence (ibid., 115-22).

5. Egoism

In the previous chapter, I described the hedonistic aspect of the theory I am defending in my thesis. In this chapter, I turn to the egoistic aspect. I start by describing egoism as a theory of ultimate reasons. Section 5.2 defends this theory against charges of incoherence or inconsistency. I then clarify, in section 5.3, the relationship between rational egoism and theories of morality.

5.1 Rational egoism

In this thesis, I am interested in egoism as a normative theory. I argue that people ought to behave as egoism prescribes, not that they already do. Psychological egoism has been convincingly criticized by a number of philosophers (see e.g. Butler 1726, Blackburn 1998, 5.2-5.5; Feinberg 2005), and I mention it only to put it aside. It is simply not true that human beings are never motivated by anything other than self-interest.

There are many versions of normative egoism, and I will be discussing the most important ones in this chapter. I start with describing the kind of egoism that I take to be correct. It is intended as a theory of ultimate reasons. When egoism is intended as a theory of reasons, it is usually called rational egoism.⁵⁵

According to rational egoism, one ought, all-things-considered, to do what best promotes one's own interests.⁵⁶ This is not because all conduct that best promotes the agent's self-interest also has additional properties that provide the agent with reason for action. Rather, maximally promoting self-interest is in itself a reason-providing characteristic. If one all-things-considered ought to perform an action, then this will be *because* – and only because – the action is in one's best interest.

The claim that one ought to promote one's interests must not be confused with any of these claims:

⁵⁵ Sometimes rational egoism is described as a theory of “rationality”, without any clarification of what rationality is or how it relates to reasons (see e.g. Shaver 2010, sect. 3).

⁵⁶ There can be non-maximizing versions of rational egoism as well. According to Michael Slote, for instance, there is only an egoistic reason to do what promotes one's interests *well enough*; there can even be a positive reason not to maximize (Slote 1989). In section 9.1, I will argue that when it comes to hedonic interests, maximizing is the thing to do. It is possible that Slote's view has some plausibility on certain non-hedonistic theories of self-interest, but I will not pursue that question here.

(1) There is a non-derivative reason for being engaged in promoting one's interests, apart from how this affects the actual fulfillment of one's interests.

(2) Being engaged in promoting one's interests is part of what is in one's interest; being an egoist automatically makes one better off.

Some examples will illustrate the difference between these claims and rational egoism. Let's say that you could choose between two lives. In the first, you are an altruist and often make sacrifices for others, but because you find yourself in very fortunate circumstances, you end up with a reasonably high degree of self-interest fulfillment. In the second life, you are a clever and consistent rational egoist and never sacrifice your interests to others, nor do you miss out on any opportunity to promote them. However, since you live under very harsh circumstances, you end up with a level of interest-fulfillment that is no higher than in the first life. According to (1), you ought nevertheless to prefer the second life. According to rational egoism, however, it does not matter which of these lives you choose. The important thing is to what extent your interests are fulfilled, and not how this comes about. In this sense, rational egoism is a consequentialist doctrine.

Imagine two persons; the first is an altruist and the second is an egoist. Apart from this, they are identical and have ended up fulfilling exactly the same set of their own interests. According to (2), the second person has then reached a higher level of overall interest-fulfillment than the first. The mere fact that he is an egoist makes his life better for him and so more desirable from the point of view of rational egoism. Similar claims could be made about being self-reliant, having power over others, or engaging in fearless struggle with one's competitors. These are, however, very controversial ideas about the content of people's interests, distinct from the claim that one's interests ought to be maximally promoted.

Neither (1) nor (2) follows from rational egoism. Some egoists may have believed in them (see Stirner 1845), but I find them very implausible and my arguments against non-hedonic intrinsic value in chapters 7-9 will apply to them. In other words, if someone thinks that being an egoist is an end in itself in either of these two senses, I believe there will be some purely psychological explanation for this belief. It is not the result of a correct perception of an ultimate reason.

Rational egoism must be distinguished from personal egoism (or individual egoism, as some authors call it). Personal egoism is a doctrine that is meant to apply only to oneself: if I am a personal egoist, I will hold that *I* ought to do what best promotes my interests, but I will

deny that this has any implications for what others have reason to do. This view is, if not incoherent, then at least implausible. It fails to satisfy what I in section 2.1 called the universalization requirement on reasons-judgments.⁵⁷

An egoist could believe that some or all other agents have no reason to act in accordance with egoism even if he accepts the universalization requirement and thus rejects personal egoism. Then he would, however, have to believe that there is something very special about himself or his interests that justifies giving them priority over the interests of others. Nietzsche may have thought that being an *übermensch* gives you a right to act egoistically, at least against those who are mere slaves (see Nietzsche 1886, sect. 260, 265; Risse 2008). Any reader who finds this way of justifying egoism promising should probably seek professional help.

Rational egoism is not based on megalomania. It claims that any agent, irrespectively of his talents and abilities, has a reason to do what best promotes his interests. Also, it satisfies the universalization requirement. If I have a reason to cheat on my taxes, then anyone who is relevantly like me and in relevantly the same situation also has a reason to cheat on his taxes. Or assume that it is in my neighbor's best interest to kill me and my family so he can steal my copy of *The Methods of Ethics*. I will naturally hope that he won't do so. But if he does, I must as a consistent rational egoist accept that he has not made any normative mistake. He does what he has most reason to do.

Rational egoists should, in other words, agree in their judgments about what ultimate reasons people have. In that sense, they are not entitled to give themselves any kind of priority. However, in another sense they clearly are. When I ask myself what *my* reasons are, it turns out that I ought to give myself absolute priority. If I have a headache, I have a reason to take an aspirin, but I never have a (non-derivative) reason to give an aspirin to anyone else. There must be something that justifies this priority. If an agent has a reason to do X and not to do Y, there must be a relevant difference between X and Y. How could there always be a relevant difference between actions that benefit me and actions that benefit others? Is not egoism simply an unjustified policy of preferential treatment, similar to racism or sexism?

This charge of arbitrariness is a standard objection to rational egoism (see J. Rachels 2002; Shafer-Landau 2003, 199; Shaver 2010, sect. 3). One way of answering it would be to

⁵⁷ A related, but perhaps even more implausible version of egoism is what Rawls calls "first person dictatorship": any agent ought, all things considered, to do what is in my best interest (Rawls 1999a, 107). As far as I know, there are no defences of this view in the literature.

point out that no other person has quite the same qualitative properties as I do. Max Stirner seems to be a representative of this view: he repeatedly requires that the egoist should conceive of himself as “unique” (Stirner 1845, see e.g. 361). This is a very weak justification of the egoist policy. First, it seems easy to imagine someone who is qualitatively identical to me, but numerically different, and then rational egoism would tell me to disregard this person’s interests as well. Second, even if every human being should turn out to have some unique properties, it is hard to see how that could make such an enormous difference when it comes to reasons for action. What would the properties in question be?

What the rational egoist ought to say, I think, is that when determining one’s reasons for action the relevant difference between oneself and other people is that one has a certain kind of access to one’s own goods that one does not have to the goods of others. No matter how many qualitative properties two people share, they cannot experience each other’s pain, nor can they possess each other’s knowledge, freedom or virtue. This is why, according to rational egoism, they have no ultimate reason to be concerned with each other’s interests.⁵⁸

I should stress that this is not supposed to constitute an argument for egoism. Any non-egoist could (and should) accept that there is a sense in which one cannot “have” other people’s goods. The question is not whether this separation between people exists, but whether and to what extent it is normatively relevant.⁵⁹ This is a difficult question, and I will be discussing it in later chapters. My present concern has just been to show that rational egoism cannot be ruled out from the outset on the formal ground that it provides no basis for the differential treatment it prescribes (see Brunton 1956, sect. 2 for a similar claim).

Rational egoism is, as we have seen, a universal theory. It is intended as an answer to the following question:

(1) What does each individual agent have most reason to do?

⁵⁸ It would seem that this way of drawing a line between one’s own interests and those of others will not be available to those egoists who think that the fulfilment of other people’s interests is part of what is in one’s own interest. (I briefly discuss this view below.) If this is true, one does not have the kind of privileged access to one’s own goods that I have talked about here. I will not consider how these egoists should make a principled distinction between their own interests and others’.

⁵⁹ This point also applies to the debate on whether utilitarians overlook “the distinction between persons” (Rawls 1999a, 24). Utilitarians do not deny that the distinction is there; they just don’t think it is intrinsically morally important.

It is important to distinguish (1) from a number of other questions, such as for instance:

(2) What decision procedure ought an agent to use?

Rational egoism is not intended as a universal decision procedure. One should not be constantly calculating what kind of conduct will best promote one's interests, since doing so would take a lot of time and effort and interfere with the fulfillment of many interests. Unless one faces unusual or important choices, it is typically better to rely on habit and various rules of thumb.

Another question to which rational egoism provides no direct answer is the following:

(3) What motives ought an agent to have?

There is a persistent tendency, both in everyday language and in the philosophical literature, to think that egoism entails that one ought to be *self-centred*, i.e. preoccupied with oneself and the fulfilment of one's own interests. However, a requirement that people must be self-centred is not a part of rational egoism (see Butler 1726; Österberg 1988, 3.1). The theory favours the motivational set that would be maximally conducive to the fulfilment of the agent's interests, and on any plausible theory of self-interest, people are generally better off having many non-egoistic desires and impulses.

A successful person by egoistic standards is not, then, a person without regard for other people, but a person who cultivates his regard for other people *to the extent that best fulfils his interests*. In order to contradict rational egoism, it is not enough to insist that there could be good reasons to have benevolent dispositions: one must be able to show that there could be good reasons to have these dispositions to a greater degree than what best promotes one's overall interests.

It might be objected that even if it should be desirable, from an egoistic point of view, to have non-egoistic desires, it will not be psychologically possible for someone who is convinced that rational egoism is true. As pointed out in section 4.2, however, our motives and impulses are not to that extent at the mercy of abstract philosophical beliefs. Being convinced that all ultimate reasons are based on self-interest does not by itself eradicate deep features of our psychology such as sympathy and conscience.

It is tempting to think that an answer to (1) will also be an answer to:

(4) In which theory of reasons does an agent have reason to believe?

If we are talking about epistemic reasons, this may be so. However, as observed in chapter 2, there is no guarantee that there are ultimate reasons for believing the correct theory of ultimate reasons. Whether one ought, according to rational egoism, to believe in rational egoism depends solely on which consequences this belief has for the fulfilment of one's interests. There are obvious benefits to being a rational egoist, but there can also be a downside. Some people may have hoped for a nobler and more inspiring account of ultimate reasons and feel depressed over the fact that nothing matters except one's own interests. Also, although one would not automatically lose one's non-egoistic sentiments, they would perhaps yield somewhat less satisfaction when they no longer are combined with a belief that they have a direct rational basis. My suspicion is that, for most people, these losses are outweighed by the gains. This is an empirical claim. Perhaps it is mistaken; perhaps it is never in anyone's interest to believe in rational egoism. This would not refute rational egoism as a theory of ultimate reasons (see Sidgwick 1907, 174-5, Österberg 1988, 6.2). We should not, at least if we are realists, reason as follows: "It seems as if only one's own interests provide ultimate reasons for action. However, it is not in one's own interests to believe in this theory of reasons for action. So other people's interests must provide ultimate reasons for action as well."

Yet another question that differs from (1) is the following:

(5) What conduct does an individual agent have reason to encourage other agents to engage in?

According to rational egoism, there is no independent reason to get other people to act in accordance with rational egoism (or, for that matter, to believe in the theory). You should only do so if that would be in your own best interest. In some cases it may be, but in others it won't.

Here is a somewhat related question:

(6) What does an agent have most reason to do if all agents will, as a matter of fact, act in the same way?

The answer is, of course, that the agent ought to act in the way that, if generalized, best promotes his interests. This means that it could be desirable for him to perform actions that

are – considered apart from their effects on the conduct of others – not egoistically desirable at all. For instance, he may have reason to stop littering or to use public transportation more often.

The condition in (6) is often not satisfied. There are many cases where an agent's choices will have little or no causal impact on how others act. We can still ask the following hypothetical question, however:

(7) If all agents were to act on the same theory of reasons, what theory of reasons ought one to prefer?

According to rational egoism, the answer to (7) depends on which kind of conduct that, if generalized, best promotes one's interests. Sometimes, such as in cases involving competitive markets, egoistic conduct may fit this bill. However, there are also cases where each individual would be better off if all the members of a group acted on a non-egoistic theory. One famous example is the prisoner's dilemma case, where two altruists will end up with a shorter prison sentence than two egoists. Another is what Garrett Hardin called the tragedy of the commons: It may be in each individual's interest to use large quantities of a certain resource, but if everybody does so, the resource will be depleted and everyone may be worse off than if all had shown some moderation (Hardin 1968).

For an egoist, the answer to (7) has no direct relevance when determining the answer to (1). The fact that the general observance of a normative principle would be conducive to the fulfillment one's interests does not give one a (non-derivative) reason to follow it when general observance cannot be secured by one's actions. If the members of a group acting on a non-egoistic theory ridicule a group of egoists for not being as well off as they could have been, the answer would be the one suggested by Parfit: "What is worse for us is that, in our group, there are no fools. Each of you has better luck. Though your irrationality is bad for you, you gain even more from the irrationality of others." (Parfit 1984, 89)

Some people seem to be puzzled by this divergence between what one has reason to do and what conduct one would find desirable if generalized. Admittedly, it could perhaps constitute a reason for opposing rational egoism for those who are anti-realists about reasons. On a realist view, however, it is more difficult to see what the objection would be. One may of course be suspicious of the divergence between the answers to (1) and (7) because of a pre-existing general conviction that one must accord some independent weight to the interests of one's group when one acts. If so, there is no independent problem with rational egoism to be

found here. Also, it is perhaps true that many moral philosophers have been working under the tacit assumption that the normative realm is such that if each of us do what he or she has most reason to do, then we will all tend to do quite well. But once we start reflecting on this presupposition, I see no reason to think that it must be correct. There is nothing incoherent or obviously implausible with the idea that if all agents individually do what they have most reason to do, then we may still end up with suboptimal results on the group level. Consider, for instance, the common sense conviction that parents ought to give special priority to their children. If all parents act on this conviction, the result will in some cases be that all children are worse off than if all parents had been neutral. Surely this does not, by itself, constitute any kind of proof that each parent has no reason to favor his children over others. What it shows is only that a parent who does favor his own child over others nevertheless could have a reason to take part in some impartial scheme for taking care of children.

This last point brings us to our final question:

(8) What legal and political system is there a reason to try to put in place?

Ideally, an egoist would want his society to be organized in a way that is maximally conducive to the fulfillment of his own specific interests. This proposal is, of course, unlikely to be supported by his fellow citizens. His efforts will, therefore, have to be concentrated on establishing a political system that is somewhat more impartial. Rational egoism puts no a priori restriction on what the best feasible system would look like – it all depends on one's interests and the willingness of others to co-operate. In particular, there is no reason to think that best system will tend to reward egoistic behavior and punish non-egoistic behavior. Hobbes based his whole moral theory on the fact that an egoist could consistently favor introducing a legal system that punishes people for doing what they, had the punishment not been introduced, would have most reason to do.

This opens up the possibility that egoists can, with the aid of legislation, solve the practical problems raised by the kind of situations mentioned above (or, better yet, prevent them from occurring in the first place). For instance, a large majority of people are presumably made better off by a decision to ban or tax deforestation and pollution. This solution will only work, of course, if agents generally find it in their interest to abide by the laws of their society. Rational egoism provides no guarantee that there is always most reason to privately follow the rules that there is most reason to publically favor. In section 5.3, I give some reasons for thinking that this will not be as serious a practical problem as one might first

suppose. But I could be too optimistic: perhaps it would be impossible to solve the major political problems of our time if everyone acted in accordance with rational egoism. This would not show that rational egoism is mistaken and that other people's interests really do provide each agent with ultimate reasons for action. It would only show that it would be a good thing for most of us if most of us act as if they do.

To sum up, then, rational egoism is to be understood as a universal theory of what individuals have reason to do, and nothing more than that.

5.2 Attempts to prove that rational egoism is mistaken

Many philosophers believe that we should reject rational egoism. Most of them do so because they are convinced that the theory makes claims about ultimate reasons that are substantively mistaken. I will discuss objections of this kind in later chapters. However, some have tried to demonstrate that rational egoism is flawed in an even more basic sense: it is, they claim, internally inconsistent or based on indefensible presuppositions. If this is true, then rational egoism cannot even get off the ground, and there would be no need to examine or defend its substantive implications. I will, therefore, discuss some objections to rational egoism of this latter kind in this section.

These objections typically make one of the following two claims:

- (1) The notion of a reason that only has authority over one individual is incoherent.

- (2) The notion of a reason that only has authority over one individual is not incoherent, but by claiming that there are such reasons one is logically committed to recognizing that there are also reasons with authority over all agents.

If any of these claims are true, rational egoism will be false. I will begin by discussing the first. An early version of it is found in G. E. Moore's *Principia Ethica*. According to rational egoism, my goods are, in a sense, good only for me. Moore thinks that this egoistic notion of good-for-me is incoherent:

when I talk of a thing as being 'my own good', all that I can mean is that something which will be exclusively mine, as my own pleasure is mine [...], is also *good absolutely*; or rather that my possession of it is *good absolutely*. The *good* of it can in no possible sense be 'private' or belong to me, any more than a thing can *exist* privately or for one person only. [...] If, therefore, it is true of *any* single man's "interest" or "happiness" that it

ought to be his sole ultimate end, this can only mean that *that* man's "interest" or "happiness" is *the sole good*, *the* Universal Good, and the only thing that anybody ought to aim at. What Egoism holds, therefore, is that *each* man's happiness is the sole good—that a number of different things are *each* of them the only good thing there is—an absolute contradiction! (PE, sect. 59, see also sect. 60)

Moore seems to think that this is an *argument* against egoism. He ends the quoted passage by saying that "no more complete and thorough refutation of any theory could be desired." However, all he actually does is to deny from the outset that the notion of a personal good, or of a reason with authority over only one individual, makes sense. He never explains why the egoist has to accept an impersonal definition of goodness or reasons.⁶⁰

It is, of course, conceivable that Moore is right in his claim that when philosophers or ordinary people call something "good", this automatically implies that any agent ought to aim at it. However, that would not show that rational egoism is incoherent. It would only show that it shouldn't be expressed in the traditional terms of goodness. The egoist would have to rely on some new terms. Perhaps Moore would insist that these terms, too, would have to be incoherent. Then it seems, however, that he would owe us an explanation of why this must be so.

In her book *The Sources of Normativity*, Christine Korsgaard tries to develop Moore's challenge to rational egoism. More precisely, she tries to show that egoism is based on the idea that reasons are "private", and this she takes to be just as untenable as the idea of a private language. She thus thinks that egoism will "die with the myth of the privacy of consciousness" (Korsgaard 1996, 136-45). Perhaps Moore had a similar idea in mind when he said that goodness "can in no possible sense be 'private' or belong to me, any more than a thing can *exist* privately or for one person only".

In order to assess this objection, we must be clear about what is meant by a reason being private. Korsgaard never defines this concept properly, but we can distinguish between three possibilities. First, it may mean *incommunicable*: if a reason is private, then there is no way of describing it to other people. This is how Richard Joyce interprets Korsgaard's argument (Joyce 2001, 125-8). He rightly points out that there is no reason to suppose that egoistic reasons have to be private in this sense. Surely nothing stops an egoist from telling people about his reasons.

⁶⁰ Jan Österberg makes the same point in Österberg 1988, 5.4. He also argues that a later, somewhat modified objection to egoism put forward by Moore fails for the same reason.

Judgments about reasons may also be private in the sense that they are not governed by what I have earlier called the requirement of universal validity. Let's say that a person claims to have a reason to cut off his leg and keep it in a plastic bag under his bed. Even if we understand this reason linguistically, we may not understand it normatively. If the person claims to have some sort of justification for his normative claim that is not, in principle, available to other agents, then his "reason" is starting to seem like a rather strange entity. It is very plausible to hold that valid claims about reasons must be universally valid, in the sense explained in section 2.1.

However, rational egoism does not violate the requirement of universal validity. It provides a criterion of what individuals ought to do that is publically justifiable in the same way as any other realist theory of reasons for action. As explained in the previous section, the fact that egoists do not have reasons to do the same things does not mean that they cannot agree on what each of them has reason to do. To be sure, one *could* combine rational egoism with a radical relativism about self-interest. Then there would be no universally valid criterion for what is in an agent's interests. However, that would mean trouble not just for rational egoism, but for any theory of reasons that gives any weight to the interests of any agent. More importantly, egoists do not tend to embrace this kind of relativism.⁶¹ Even a simple preferentialist account of self-interest attempts to provide a universally valid criterion for when an agent has done what is in his best interest.⁶²

In particular, the hedonistic version of rational egoism is not private in the present sense. Nothing stops believers in the doctrine from reaching agreement on which reasons each of them has. They will not simply accept a purported reason to cut off one's leg at face value, but insist that there is only such a reason if cutting off the leg maximizes the agent's long-term hedonic level. If it is claimed that there is some other justification for cutting it off, they conclude not that the reason claim is privately correct, but that it is publicly false.

Finally, a reason may be private in the sense that it only has independent authority over one agent. This is the kind of privacy that Korsgaard needs to rule out in order to show that egoism is mistaken. However, it does not seem that this kind of privacy is metaphysically or linguistically objectionable, at least not in any way that resembles the strangeness of a private

⁶¹ An exception is Max Stirner who holds that the egoist should create his own truths (Stirner 1845, 342-54). This includes, presumably, the truth about what is in his interest.

⁶² Joshua Gert makes the same point in his criticism of Korsgaard's argument (Gert 2002, 312-3).

language. It does not imply that a reason is an entity that can only be identified or evaluated by a single person.

Korsgaard confounds the second and the third sense of privacy in her argument. The following passage is typical:

Consider an exchange of reasons. A student comes to your office door and says: 'I need to talk to you. Are you free now?' and you say 'No, I've got to finish this letter right now, and then I've got to go home. Could you possibly come around tomorrow, say about three?' And your student says 'Yes, that will be fine. I'll see you tomorrow at three, then.'

What is happening here? On my view, the two of you are reasoning together, to arrive at a decision, a single shared decision, about what to do. And I take that to be the natural view. But if egoism is true, and reasons cannot be shared, then that is not what is happening. Instead, each of you backs into the privacy of his practical consciousness, reviews his own reasons, comes up with a decision, and then re-emerges to announce the result to each other. And the process stops when the results happen to coincide, and the agents know it, because of the announcements they have made to each other. (Korsgaard 1996, 141)

As pointed out above, egoism is not intended as a universal decision procedure, and this clearly seems to be a situation where explicit egoistic reasoning would be out of place. However, let us disregard that and assume that you are both deliberating in terms of self-interest. It is true that this conversation would then not be a search for a single reason that has authority over both you and the student. But that does not mean that you must deliberate in the privacy of your own practical consciousness, in the sense that all you can do is announce various times you would be willing to have a meeting, let the student do the same and then just hope for a match sooner or later. That would indeed be a strange picture of everyday appointment-making. Rather, the two of you will be engaged in a common enterprise in the sense that you search for a time that is egoistically acceptable to both of you. It is not as if you are denied epistemological access to each others' reasons. You will, for instance, realize that the student is more likely to be interested in supervision tomorrow than in two weeks. If he claims that he has an extremely important question about the paper he is working on, you may reconsider whether you ought to go home, since there is now (assuming it is in your interest to do a good job) a stronger reason to work late. On the other hand, you may reassure the student that there is still plenty of time before his paper is due, and that students tend to greatly overestimate the importance of certain details in their papers.

More generally, the egoist outlook allows that other people can tell you what to do, in the sense that they can explain the correct theory of reasons to you or inform you of normatively

relevant facts, thereby getting you to change your beliefs about what you ought to do. Denying that other people can obligate us in this indirect way would perhaps be incoherent or solipsistic. But denying that other people can obligate us directly, by providing us ultimate reasons to be concerned with their interests, is not.

Moore and Korsgaard claim that it does not make sense to talk about ultimate reasons with authority over only one individual. Others have made a somewhat more moderate objection: even if it makes sense to talk about egoistic ultimate reasons, it is inconsistent to hold that that *all* ultimate reasons are egoistic, since one is, by believing in ultimate reasons of any kind, logically committed to a belief in certain non-egoistic ultimate reasons.

Kant is the most well-known representative of this strategy. He appears to give two different arguments for the claim that any rational being must take the interests of other into account when deciding what to do. The first starts from the postulate that an ultimate reason, or, in Kant's terminology, a categorical imperative, must "conform to universal law" (Kant 1785, 421). What does this mean? One possibility is that Kant wants to embrace what I have earlier called the universalization requirement on reasons-judgments. In Kant-inspired terms, it goes like this: "Act only on the maxim whereby you can at the same time acknowledge that any agent would have reason to act on that maxim." For instance, if you think you ought, in your situation, to make a false promise to in order to obtain money, then you are committed to holding that any agent so situated ought to do so.

Kant, however, goes on to derive an importantly different principle: "Act only on that maxim whereby you can at the same time will that it become a universal law" (ibid.). This principle requires not only that you acknowledge that others would have reasons to act on your maxim, but that you can *will* that they do so. And this means, as Kant points out, that it rules out certain kinds of behavior that could very well be approved by rational egoism, such as for instance making false promises in order to obtain money.

Kant's justification of his categorical imperative seems to amount to little more than a conflation of what I above called question (1) and (7). Kant wants to show that any rational agent is bound by non-egoistic principles, but what he offers in defense of this claim is only the observation that any agent ought to prefer all agents adhering to a non-egoistic principle over all agents adhering to rational egoism. Maybe so, but why is that relevant when my choices will not, in fact, determine what principles other people adhere to? I do not want to deny that Kant's categorical imperative could be correct as a substantive claim. Perhaps there is even a way of deriving it from some more basic normative notion, or some way of showing

that question (1) and (7) are equivalent. The important point for now is that Kant has not in fact done so.⁶³

Kant's other argument is, like the first one, very brief and not particularly clearly formulated. It runs like this:

The human being necessarily conceives of his own [rational nature as an end in itself]; so far, then, this is a subjective principle of human actions. But every other rational being regards its existence similarly, just on the same principle that holds for me; so that it is at the same time an objective principle from which as a supreme practical law all laws of the will must be capable of being deduced. Accordingly the practical imperative will be as follows: *So act as to treat humanity, whether in your own person or in that of any other, in every case at the same time as an end, never as a means only.* (ibid., 429)

Let us, for the sake of argument, accept that a given agent must conceive of his rational nature as an end in itself. This means, of course, that any other agent must conceive of *his* rational nature as an end in itself. But what Kant immediately goes on to say is something much stronger: any given human being must conceive of the rational nature of *all* human beings as an end in itself. This contradicts rational egoism, but it is hard to see how it could follow from Kant's premises. How does he get from the claim that one's own rational nature is a source of ultimate reasons for oneself to the thesis that anyone's rational nature is a source of ultimate reasons for anyone?

Allen W. Wood suggests that Kant can solve this problem by saying that "the reasons in question have to be not merely *motivating* reasons, but also *justifying* reasons, and that justifying reasons always require some universal grounding." (Wood 1999, 128) If "universal grounding" just means meeting the requirement of universal validity, then this claim would be plausible, but irrelevant, since egoistic reasons are also universally grounded in this sense. If, in the other hand, it means having universal authority, then it amounts to nothing more than an assertion of the view that was supposed to be argued for. In any case, we find no explanation of why accepting the existence of egoistic ultimate reasons must lead us to accept the existence of non-egoistic ultimate reasons as well.

⁶³ Allen W. Wood, who is a prominent contemporary Kantian, also denies that Kant is successful in deriving his principle from mere idea of a categorical imperative (Wood 1999, 78-82).

A similar, but more recent and carefully developed argument can be found in Alan Gewirth's book *Reason and Morality*. As we will see, it suffers from the same basic problem. This is the argument:

A prospective purposive agent (PPA) claims (by definition)

(1) I do (or intend to do) X for some purpose E.

By virtue of making this claim, the PPA rationally must consider that

(2) E is good.

(3) My freedom and well-being are generically necessary conditions of my agency.

(4) My freedom and well-being are necessary goods.

By virtue of having to accept (4), the PPA must accept

(5) I have a claim right to my freedom and well-being.

From (1)-(5) it follows that the PPA must consider that

(6) I am a PPA \rightarrow I must consider that I have a claim right to my freedom and well-being.

From (6), combined with the logical principle of universalizability, it follows that the PPA must consider that

(7) Another person is a PPA \rightarrow the person has a claim right to his freedom and well-being.

Since every PPA is required to reason in this way, it follows that

(8) Every PPA must accept that every PPA has a claim right to his freedom and well-being.⁶⁴

If this argument is sound, the only way to deny that all agents have rights to freedom and well-being is to deny that one is an agent. Since rational egoism is addressed to agents, that would, in turn, mean that rational egoism is inconsistent.

I will not discuss the first four stages in this argument. At first sight, it may seem false that (5) follows from (4): the fact that something is a good – even a necessary good – for an agent does not entail that the agent has a right to it. However, Gewirth makes it clear that the right

⁶⁴ My formulation of it is adopted (in a somewhat modified form) from Beyleveld 1991, ch. 2.

in (5) is only “prudential” (Gewirth 1978, 2.8-2.10). It is, at this stage, not claimed that it has independent authority over other agents. In fact, (5) is to be understood as saying nothing more than “it would be conducive to my prudential purposes that other people not interfere with my doing X” (ibid., 94).

This qualification ensures that (5) does indeed follow from (4). However, the consequence is that the right in (6)-(8) must also be nothing over and above a prudential right. What the argument shows is thus only that any agent is committed to holding that any agent has a prudential right to his freedom and well-being, which again only means that it is conducive to the prudential purposes of any agent that others not interfere with his actions. Clearly this last claim is no contradiction of rational egoism.

Many authors have made similar objections to Gewirth’s argument. Korsgaard, for instance, says that consistency “can force me to grant that your humanity is normative for you just as mine is normative for me. It can force me to acknowledge that your desires have the status of reasons for you, in exactly the same way that mine do for me. But it does not force me to share in your reasons, or make *your* humanity normative for me.” (Korsgaard 1996, 133-4; see also Hare 1984; Kalin 1984, 141-4; B. Williams 1985, ch. 4) Gewirth is not convinced by these objections (Gewirth 1984, 207-12). Dereck Beyleveld has even written a book-length defense of his argument (Beyleveld 1991; see in particular 257-281). I find that their reformulations and counter-objections do not bridge the gap in the argument, but only serve to make it less obvious. It would take too much time to give a thorough defense of this claim. I will, therefore, end this section by just restating my (relatively widely shared) conviction that there is no way of showing that one is, by conceiving of oneself as an agent who acts on reasons, logically committed to denying rational egoism. In other words, if there turns out to be something wrong with rational egoism, it will be not that it is inconsistent, but that it is substantively mistaken.⁶⁵

⁶⁵ I will not discuss Thomas Nagel’s claim in *The Possibility of Altruism* (Nagel 1970) that rational egoism presupposes solipsism, since Nagel later admits that it is mistaken (Nagel 1986, 159; see also Sturgeon 1974). I postpone criticism of his anti-egoistic argument in *The View from Nowhere* to chapter 9, since this argument is not meant to show that rational egoism is in any way inconsistent or incoherent, but only that it is in normative error.

5.3 Rational egoism and morality

Assume that rational egoism is correct. How would that affect our beliefs about what is morally right and wrong? The answer to this question depends on one's understanding of moral concepts.⁶⁶ There seems to be three main alternatives in the literature.

First, some moral philosophers think that it makes sense to talk about morality without independent authority. They see moral theories not as theories about what one all-things-considered ought to do, but rather as (something along the lines of) proposed standards for impartial evaluation of conduct. It is one question, they think, what the correct moral standard is, and another question whether there is an ultimate reason to act in accordance with it. The moral standard is thus defined by its content, or domain, rather than by its authority.

Of course these authors are not denying that morality could turn out to have independent, or even supreme, authority. The important point is that this would be a substantive normative discovery, and not just a definitional truth. If it turns out that there is no such moral authority, that would not in any way tend to show that there is no such thing as morality. In fact, we find this understanding of moral theories in several prominent recent defences of moral realism.

For instance, Russ Shafer-Landau writes:

Do [moral] facts necessarily supply us with reasons for action? It is important to see that realism *per se* is neutral on this question. Whether moral facts invariably supply reasons for action depends not on realism alone, but very importantly on which theory of practical reason one adopts. (165) Realists who reject rationalism should simply insist that morality is only contingently reason-giving; moral demands supply reasons for action only when they align, for instance, with one's desires or interests. This is a perfectly consistent view. (Shafer-Landau 2003, 192)

Similar definitions of moral realism are given by Peter Railton and David Brink (Railton 1986, sect. V; Brink 1989, ch. 3). Shafer-Landau goes on to argue for the substantive claim that moral facts do in fact always provide us with reasons for action (op. cit., ch. 8).⁶⁷ Brink, on the other hand, is skeptical of the independent authority of morality, and seems to be

⁶⁶ Some authors distinguish between moral and ethical concepts. One might, for instance, hold that any normative theory, including rational egoism, is an ethical theory, while only some normative theories are moral theories. (For a more complicated proposal, see B. Williams 1985, ch. 1.) However, in the literature on egoism, ethical and moral terms are typically used interchangeably, and I will follow this usage in my thesis.

⁶⁷ Somewhat confusingly, denying moral rationalism, which he a few paragraphs earlier considered "perfectly consistent", then seems to him "a conceptual error" (ibid., 192-3). Presumably what he means to say is just that moral rationalism is intuitively plausible and accepted by many agents.

leaning towards rational egoism (op. cit., 3.12, 8.6). But they all agree that it is an open question whether morality has independent reason-generating power.⁶⁸

Contemporary utilitarians tend to conceive of their moral theory in a similar fashion. Peter Singer, for instance, thinks that the principle of utility is, objectively speaking, the correct standard for impartial evaluation of conduct. However, he does not think that it is, in itself, a source of ultimate reasons (Singer 1993, 314-22). In order to show that it is generally “rational to act morally”, he finds it necessary to show that it is generally “in our long-term interest to do so” (ibid., 322-35; see also Hare 1981, ch. 11; Brink is also a utilitarian).⁶⁹

If we end up thinking that morality does not have independent authority, we face a certain complication in moral discourse. Since we cannot use the independent authority to fix the reference of moral terms, it is no longer obvious that there is such a thing as *the* moral standard. Are we, to name a few possibilities, looking for the standard of a benevolent and omniscient observer, the standard that would be chosen behind a veil of ignorance, the standard that an epistemologically rational person would want to see implemented in his society, the standard implicit in our moral concepts, or a standard that no-one could reasonably reject? Once we have specified the moral question in either of these directions, it may have a determinate and objective answer, but how do we agree on the question?

We may perhaps be able to show that some of these questions are more interesting than others. Also, it may turn out that the most reasonable formulations of the moral question have very similar answers. I will not pursue these issues any further, since it does not seem as if the plausibility of rational egoism depends on it. The important point for my purposes is that rational egoism, on this understanding of morality, has no implications whatsoever about the moral rightness and wrongness of actions. It is compatible with any view on what the correct moral standard is, be it consequentialist or deontological.

The acceptance of a *hedonistic* version of rational egoism would, however, put some limitations on the content of moral theories. There should not be any non-derivative duties to bring about non-hedonic goods, such as knowledge, promise-keeping or autonomy. (I suppose combining such duties with hedonism about ultimate reasons would not be inconsistent; it just

⁶⁸ Christine Korsgaard also thinks it makes sense to question whether morality is an independent source of authoritative reasons for action, or, in her terms, whether it is “normative”. At the same time, she interprets moral realism as an attempt to demonstrate the authority as well as the universal validity of moral standards (Korsgaard 1996, see especially lecture 1).

⁶⁹ Singer never explicitly endorses rational egoism. It is possible, but unlikely, that he holds that there are ultimate reasons that are neither egoistic nor moral.

wouldn't be very plausible.) Hedonistic utilitarianism would fulfill this requirement, and many hedonists have indeed been utilitarians. However, it should be pointed out that this is not the only possible moral standard for a supporter of hedonistic rational egoism. One could, to name a few examples, hold that the amount of pleasure experienced by people should be as equal as possible, that those who have suffered most should be given some priority, that everyone has an inalienable right to a certain minimum hedonic level, or that there is a limit on permissible interpersonal trade-offs. It seems to me that these alternatives to utilitarianism are difficult to defend except by appeal to the kind of normative intuitions I warn against in later chapters. My personal conviction is, therefore, that classical hedonistic utilitarianism is the most plausible answer to the most reasonable formulations of the moral question. However, I will not provide any further arguments for this claim. In the remainder of this thesis, I will not rely on or presuppose any specific theory of the content of moral reasons.

Even if a requirement of independent authority should be implicit in our pre-theoretical notion of morality, it might be claimed that this requirement is not so essential that we have to abandon all moral discourse if it cannot be satisfied. We could accept Singer and Brink's account of moral concepts, not as an accurate report of how the concepts actually are used by philosophers and in everyday language, but as a revisionist doctrine made necessary by philosophical reflection. This is in fact the response to rational egoism that seems most reasonable to me. However, some philosophers evidently have a different pre-theoretical understanding of morality, according to which independent authority *is* an essential feature, and I see no basis for claiming that mine is the "real" pre-theoretical understanding. I will, therefore, also consider the consequences of rational egoism for morality if we assume that morality by definition must possess independent authority.

Some philosophers hold that theories of morality simply *are* theories of reasons: whatever we all-things-considered ought to do is also what is morally right to do. They thus define the moral standard only in terms of its supreme authority, and put no a priori restrictions on its content. This seems to be Sidgwick's way of thinking. For him, any procedure that tells us what is "reasonable" to do or what our "ultimate end" should be is an ethical method (Sidgwick 1907, book I, ch. I, III). He thus considers it a conceptual truth that it is "right to act reasonably" (ibid., 343-5; see also Rachels 2009, sect. 3; Tännsjö 2010, ch. 3).

If this is what we mean by "morally right", then rational egoism will of course be a moral theory: conduct will be morally right if and only if it is in the agent's best interest. We might as well call the doctrine ethical egoism. However, it must be kept in mind that this is only so

as long as we by “morally right” mean nothing more than “supported by the strongest reasons for action”. In particular, we must be careful not to let any notion of impartiality along the lines outlined above be implicit in moral terms.

Jesse Kalin seems to understand the ethical egoism he defends in this way. He notes that egoism lacks certain features had by most traditional moralities and then goes on to claim that it nevertheless should be considered a moral theory: “The egoist’s basic question is “What ought I to do; what is most reasonable for me to do?”⁷⁰ This question seems to me a moral question through and through, and any coherent answer to it thereby deserves to be regarded as a moral theory.” (Kalin 1970, 86; see also Rand 1964, 13)

Many contemporary philosophers – actually most of them, I suspect – hold a view on moral terms that constitutes a kind of middle position between the two outlined so far. On the one hand, they do think that a claim to independent authority is implicit in moral concepts. We should not call something a genuine moral standard unless it has the ability to provide ultimate reasons. On the other hand, they do not think that moral reasons are the only reasons which can have independent authority. For instance, prudential or aesthetic reasons may also partly determine what one all-things-considered ought to do – even if they do not have anything to do with impartially desirable conduct and so fail to be moral reasons.

Morality is thus, on this view, defined both in terms of its content and its authority (see Finlay 2008, 139-43). A moral reason must both have a certain kind of impartial content and a certain authority. As pointed out, there is no general agreement on precisely how the content or domain of moral reasons should be delimited. Also the supporters of the present view differ on how strong the authority is supposed to be. Some insist that it must be overriding: we always have most reason to do what we are morally required to do.⁷¹ According to Kurt Baier, for instance, being overriding is morality’s very *raison d’être* (Baier 1958, 309; see also Darwall 2006, 97-9). Alan Gewirth begins *Reason and Morality* by making the same claim:

Persons guide their lives in many different ways. Among the various goals, rules, habits, ideals, and institutions that figure more or less implicitly in such guidance, morality has a unique status. For it purports to set, for everyone’s conduct, requirements that take precedence over all other modes of guiding action, including even the self-interest of the persons to whom it is addressed. (Gewirth 1978, 1)

⁷⁰ Kalin is not a personal egoist, so the question he really has in mind is presumably: “what ought *one* to do?”

⁷¹ This formulation of the overridingness thesis is explained and defended in Stroud 1998, 170-5.

A somewhat more moderate conception of moral philosophy is the idea that although moral reasons are not systematically stronger than non-moral reasons, they must at least have *some* independent authority. Michael Smith defends this view when he argues that it is a conceptual truth that “if it is right for agents to ϕ in circumstances C, then there is a reason⁷² for agents to ϕ in circumstances C.” (M. Smith 1994, 62, see also the rest of ch. 3; Joyce 2001, ch. 2; Cuneo 2007, 36-9)

If we adopt this third understanding of moral concepts, accepting rational egoism leaves us with two options. The first is to accept ethical egoism: egoism is not only the correct theory of reasons for action, but also the correct theory of morality. I will discuss this doctrine in some detail, since doing so will shed some light on rational egoism as well. The second option, which I will return to below, is to adopt moral nihilism.

Given rational egoism, ethical egoism obviously satisfies the requirement of independent authority. However, according to the present understanding of morality, there are also further restrictions on what counts as a moral standard. As we will see, egoism fails to satisfy many of the restrictions that are commonly found in the literature.

One principle that ethical egoism cannot accommodate is the following:

(1) If one agent’s strongest moral duty is to pursue X, then all agents’ strongest moral duty is to pursue X.

If we combine (1) with ethical egoism, we get the result that my strongest moral duty will be to pursue my own interest, that my strongest moral duty will be to pursue your interests, and so on for all existing agents. This is the “absolute contradiction” that Moore thought he had identified in egoism. Brian Medlin levels a similar criticism towards ethical egoism in his paper “Ultimate Principles and Ethical Egoism” (Medlin 1957). But an ethical egoist could simply deny (1), as does Kalin (Kalin 1970, sect. IV). So would perfectionists and many deontologists, I suppose.

However, there is a much more reasonable formal requirement on rightness that the ethical egoist also would have to reject:

⁷² It is clear from the context that Smith is here talking about a reason with independent authority.

(2) If an agent is morally required to do X as long as no-one interferes with his doing X, then other agents have at least some moral reason not to interfere with the agent's doing X.

According to the ethical egoism, the rightness or wrongness of an action is determined solely by how it affects the agent's own interests, and it is of no intrinsic importance whether it prevents others from doing what would otherwise be their moral duty. Egoist morality is thus not in any sense a cooperative venture. This raises doubts about whether it, on the present understanding of moral concepts, should be described as a form of morality at all.⁷³

Notice that a corresponding principle about reasons in general cannot with the same plausibility be said to embody a conceptual truth:

(3) If an agent has most reason to do X as long as no-one interferes with his doing X, then other agents have at least some reason not to interfere with the agent's doing X.

Some may think that the denial of (3) makes egoism less plausible as a theory of ultimate reasons, but it hardly tends to show that egoism fails to even be a theory of ultimate reasons in the first place.

When people have conflicting interests, we often expect there to be a possible outcome of the conflict that can be acknowledged to be morally desirable by all the parties involved. For instance, if A and B come across a pile of apples and they both want as many apples as possible, there should be a certain distribution of the apples that both A and B could agree on as morally appropriate. Similarly, we tend to proceed on the assumption that moral questions such as "should abortion be allowed?" or "should we assist the poor?" will receive the same answer from all people who are thinking clearly. Ethical egoism, however, cannot give us any interpersonally valid ranking of possible outcomes or states of affairs. A given course of action will be in the interest of some people, but not others – and that is all the theory has to say about the matter. Kurt Baier thinks this means that ethical egoism fails to even qualify as a candidate for the correct moral theory (Baier 1993, 201-3). Some authors have defended

⁷³ Alan Gewirth's charge that ethical egoism issues "incompatible directives" (Gewirth 1978, 2.11) is based on tacitly assuming something like (2).

Another principle that the ethical egoist, for similar reasons, will have to reject is that it cannot be morally wrong to openly promulgate the true moral theory. (Medlin makes a further inconsistency charge based on combining ethical egoism with this principle (Medlin 1957).) That may not be a serious problem, however. The principle is also incompatible with other forms of consequentialism, such as utilitarianism.

ethical egoism from this charge by suggesting that it was a mistake to expect that a moral theory should be able to provide this kind of interpersonally valid ranking (see Kalin 1970, 70; Regis Jr. 1980, 57-9). That may be true. But to the extent that we did expect it, egoism's claim to be considered a moral theory is correspondingly weakened.

As explained above, egoism is sometimes self-defeating on the collective level: a group of egoists may do worse, in terms of interest-fulfilment, than a group of, say, utilitarians or Kantians. Although we may accept that we cannot require theories of reasons to be successful on the collective level, we would certainly expect a theory of morality to be so. Again, the ethical egoist could insist that we were mistaken in forming this expectation. And again, that would not change the fact that we did so.⁷⁴

There is a possible modification of ethical egoism that would solve all the problems mentioned so far: We could say that an action is right if and only if it maximizes the amount of (successful) egoistic behaviour (or self-reliance, or struggle for power) in the universe. This principle is not incompatible with (2) above, it provides an impartial ranking of claims, and it is not self-defeating on the collective level. Notice, however, that this standard will demand actions that are not in each agent's interest. (I might, for instance, be morally required to give away my wealth, so that a large group of poor people could fight over it.) Therefore it cannot, if rational egoism is true, have independent authority. And the idea that moral standards must have independent authority was the very reason for considering ethical egoism in the first place.⁷⁵

Until now I have only considered objections to ethical egoism that are relatively formal: they do not directly concern the practical implications of the doctrine. One may, however, also hold that there are limitations on what practical implications that could reasonably be referred to as moral (see e.g. M. Smith 1994, p. 40). If so, ethical egoism is likely to be in trouble. Imagine, for instance, a case where you are bored and the best available way of improving your hedonic level would be to set an innocent person on fire (you have always wondered what it looks like). As long as we make sure there are no bad long-term consequences (you

⁷⁴ Jan Österberg thinks that its self-defeatingness on the collective level is sufficient to discredit ethical egoism. The same is true, he argues, for its failure to be jointly satisfiable (in some cases it is logically impossible that all agents can act rightly on the egoistic criterion) and to secure Pareto-optimality (Österberg 1988, ch. 7).

⁷⁵ If one thinks that morality does not need to have independent authority, one may of course propose this impartial version of egoism as the correct moral standard. However, the objections to ethical egoism mentioned below would still apply.

aren't caught and you feel no remorse or regret etc.), ethical egoism implies that you are then morally required to do so. You would act wrongly if you sit down and watch TV instead.⁷⁶

It is doubtful whether a standard that demands actions of this kind can be recognized as a moral standard at all. A certain minimum regard for the interests of other people seems to be implicit in moral concepts, as they are normally used both by philosophers and in everyday life.

Consider the corresponding claim about reasons, that there would be most reason to put the innocent person on fire. This claim may strike some readers as just as implausible as the claim that it is morally right. (I will return to these kinds of cases later and explain why I think that appearances are misleading.) But at least it clearly makes sense as a suggestion about what you have most reason to do. We do not have to reject or revise the concept of a reason if rational egoism should turn out to be correct. In this respect, I suggest, rational egoism may differ from ethical egoism.

Some ethical egoists openly admit that their theory has, in many cases, rather extreme implications for the treatment of other people (see e.g. G. Williams 1948). However, many ethical egoists want to avoid making this concession. (Robert G. Olson even thinks that the general adoption of egoism as a moral code would have "altogether wholesome" social consequences (R. Olson 1961).) One way of doing so is to appeal to perfectionist or other "noble" accounts of self-interest. Plato and Aristotle are well-known proponents of this strategy. My arguments against non-hedonic value in later chapters will apply to these accounts.⁷⁷

Another possibility is to hold that the interests of other people are somehow fused with one's own. On David Brink's objective list theory of welfare, "one is better off when another's welfare is enhanced, and especially when one enhances another's welfare." (Brink 1989, 242)⁷⁸ Lester Hunt, in a presentation of Ayn Rand's egoism, claims that "one's values include, as a part of them, the good of certain other people." (Hunt 1999, sect. IV, see also sect. VI) Rand herself held that that it cannot be in anyone's interest to hurt other people; indeed, true interests do not ever conflict (Rand 1964, ch. 4; see also T. Smith 2000, 174-86).

⁷⁶ James Rachels claims, relying on a similar example, that we should reject ethical egoism simply because it is "wicked" (J. Rachels 1974, sect. 2).

⁷⁷ For an argument that a "high-minded" egoist (an egoist who values virtue for its own sake) must be a prig, that is, value his own character more than anything else, and so be incapable of having a good character at all, see Lemos 1984. For a response to the charge that virtue ethics is ultimately egoistic, see Annas 2008.

⁷⁸ Brink subscribes to rational egoism, not ethical.

These egoists want to be able to say that it is (usually) not right to perform acts that most people would consider clearly immoral. This can be done by making two different claims:

- (1) Other people's interests provide one with ultimate reasons.
- (2) Other people's interests do not provide one with ultimate reasons, but what happens to other people can *by itself* influence how well off one is, and there is an ultimate reason for being well off.

I cannot see any good reason for defending the egoistic (2) instead of (1) here. Perhaps egoists think it is mysterious how non-egoistic reasons can have independent authority, but it would be just as mysterious how what happens to other people can directly determine how well off one is. If my arguments later in this thesis are sound and (1) is implausible, so is (2).

One can also try to avoid the bothersome implications of ethical egoism by advocating a more indirect application of the doctrine. It may, for instance, be in the long term interest of human beings to strictly abide by certain non-egoistic secondary principles. David Gauthier is a representative of this strategy. (So is Hobbes, perhaps.) Gauthier thinks that an agent is more likely to achieve a high level of interest-fulfilment if he gets rid of the disposition to be a maximizing egoist, and instead becomes a "constrained maximizer" (Gauthier 1986). A constrained maximizer is willing to sacrifice his interests to a certain degree, provided that others do the same. (Gauthier stresses that he is talking about genuine sacrifices here; the constrained maximizer is not just a maximizer with a long-term perspective (*ibid.*, 169-70).) One may of course doubt whether a completely rigid disposition to be a constrained maximizer really is the ideal psychological make-up, as seen from an egoistic point of view. (I will describe what I take to be a more egoistically desirable commitment to impartial principles below.) A more serious problem is that Gauthier can only justify constrained maximization when mutually beneficial co-operation with other agents is possible. His contractualism provides no direct moral duties towards children, the mentally challenged or cultures with inferior technologies (*ibid.*, 17-9, 231-2). If someone were to find himself in circumstances where co-operation would be of no use to him, he would be morally free to set fire to whomever he likes.

Perhaps it would be possible, by combining all the strategies mentioned above, to produce a version of ethical egoism that has no embarrassing practical implications. We would never have to admit that, say, a child molester acted as he was morally required to. It may nevertheless be true that these pronouncements would be justified in the wrong way for

egoism to be a plausible *moral* theory. What we would be saying, in effect, is that the sexual abuse of children is not permitted *because it hurts the perpetrator*. Even if it does, this isn't typically what we have in mind when we say that it is morally wrong. Many philosophers would say that the notion of morality includes an essential element of other-regardingness, not only on the level of prescriptions for action, but also on the level of their justification.

The egoist standard has, I believe, supreme authority, but as we have seen there seems to be little else that could justify its being labeled a moral standard. If one believes that being a moral reason is something distinct from or more than being an ultimate reason, then one should not, in other words, accept ethical egoism. At this point, revising one's moral concepts so that they no longer contain a claim to independent authority may be starting to seem like a good idea after all. Or one might perhaps find it more plausible to agree with Sidgwick and Kalin that any ultimate reason may be considered a moral reason. One can of course also continue to insist that moral reasons, if they are to be genuinely moral reasons, must both be authoritative and have a certain kind of content. In that case, rational egoism is most plausibly taken to imply moral nihilism. If one only has ultimate reasons to promote one's self-interest, then all first-order moral judgments will be false (or perhaps neither true nor false).

It must be kept in mind that this would be a limited kind of nihilism. It is not a nihilism about ultimate reasons. Nor is it a denial that there can be correct impartial standards for evaluating conduct. It is only a denial that there is such a thing as morality, understood in a particular (though admittedly not uncommon) way.⁷⁹

⁷⁹ A lack of authority is of course not the only thing that may lead one to conclude that a principle is not genuinely moral. Some philosophers believe that any sincere moral judgment must be accompanied by some motivation to act in accordance with it, but I see no reason for supposing that a rational egoist just *has to* have a desire (however weak) to do what is desirable from an impartial point of view. That would be another possible reason for calling the impartial principles I have discussed "moral" and not moral.

Also, we may expect that there should be some reliable connection between our conscience and moral principles. If hedonism is correct, this is not the case. We do feel that is wrong to inflict suffering on others, but many of our intuitive moral reactions have little or nothing to do with pain, and hardly any of them have anything to do with pleasure (though some of us may have a vague feeling that pleasure is sinful, since we often have to resist it in order to do what is right). We do not *feel* morally obliged to promote pleasure. Thus one might think that a hedonist theory of morality is not really about what we took morality to be at all. This was Seneca's response to the hedonists: "Why do you couple things that are unlike, nay, even opposites? Virtue is something lofty, exalted, unconquerable, and unwearied; pleasure is something lowly, servile, weak, and perishable, whose haunt and abode are the brothel and the tavern." (*Moral Essays*, vol. 2, 113-5)

The aim of this section has been to explain the relationship between rational egoism and theories of morality. This has been a more extensive task than one might first expect, since philosophers have differing conceptions of what a moral theory is. Some will want to say that egoism about ultimate reasons is itself a moral view, others that it is a theory with no moral implications whatsoever, and yet others that it implies moral nihilism.

Whichever normative vocabulary one prefers, it is important to make sure that it leaves room for rational egoism. Writers sometimes dismiss egoism as a normative theory because their terminology does not allow them to express the doctrine as concerned with the question to which it is a plausible and interesting answer. There is a tendency in much of the literature on the subject to treat personal egoism and ethical egoism as the only possible versions of normative egoism. Either egoism is only meant to apply to the person embracing it, and so becomes a doctrine without philosophical significance, or it is supposed to constitute a straightforwardly moral theory. If a moral theory is understood as nothing over and above a theory of ultimate reasons, this is of course not a problem. However, many of those who have written on egoism conceive of moral standards as standards that have a particular kind of content or are able to play a particular kind of practical role. Jan Österberg's *Self and Others*, which is (otherwise) one of the best book-length treatments of egoism available, provides a good example. Österberg reasonably rejects personal egoism and then takes this to imply that egoism must be able to fulfill requirements that we typically make of Kantianism, utilitarianism and other impartial moralities, such as that it should not be collectively self-defeating (Österberg 1988, 36, 7.3; see also Medlin 1957, sect. 1; van Ingen 1994, ch. 3). This is of course a test that egoism fails to pass, and so the reader is left with the impression that no form of normative egoism is tenable. Rational egoism, as defined in this thesis, is never considered. However, it is, as far as I know, the only plausible version of normative egoism there is.

6. The Implications of Hedonistic Egoism

I turn now to the practical implications of the hedonistic version of rational egoism. After describing some possible views on what the proper object of egoistic concern is (section 6.1), I discuss in what sense and to what extent an egoist would be committed to morality (section 6.2) and friendship (section 6.3).

I will argue that the way of life recommended by hedonistic egoism is not very different from the one chosen by most actual people, and certainly not as repulsive and anti-social as some authors have imagined. It should be noted from the outset that this understanding of the implications of hedonistic egoism plays no direct role in my subsequent argument for the position. There are, however, other reasons for devoting a chapter to this issue. First, if hedonistic egoism is true, it directly concerns how one ought to live. Second, even those philosophers who think that egoism is false will typically acknowledge that reasons of self-interest nevertheless have considerable force. Knowing the extent to which these reasons support a commitment to the well-being or integrity of other people will then be important for determining how strong a commitment of this kind one all-things-considered ought to have. Finally, some readers may not share the highly critical attitude towards normative intuitions defended in later chapters, and so think that a given version of rational egoism will be considerably more plausible if it can be shown that it generally does not have particularly counterintuitive implications about real-life cases.

6.1 The object of egoistic concern

What egoism tells us to do obviously depend on empirical facts about the consequences of various actions. Somewhat less obviously, perhaps, it also depends on issues concerning the nature of persons. It is clear that according to rational egoism I ought to be concerned with Ivar Labukt's present interests. But why and to what extent should I be concerned with the interests of the person called Ivar Labukt 20 years from now? How do egoistic reasons extend into the future? What exactly is the object of egoistic concern?

This is a difficult question. One possibility is that I should be completely neutral between all the parts of Ivar Labukt's life. Perhaps I and Ivar Labukt of the year 2030 are one and the same Cartesian ego persisting through time. If so, it seems plausible to say that I should be just as concerned with my 2030-interests as I am with my 2010-interests. There would be no egoistic reason for according any of Ivar Labukt's interests more or less weight than the

others. Roger Crisp suggests a view that, though theoretically different, has the same implications (at least under all normal circumstances). According to Crisp, what grounds rational egoistic concern is “continuity of capacity for consciousness”. This continuity is found throughout human lives, and so one should be strictly neutral between all the different parts of one’s life (Crisp 2006, 5.1).

Another possibility is that my future interests should be given somewhat less weight than my present. Derek Parfit defends a version of this idea (Parfit 1984, ch. 10-13). Parfit is a reductionist about personal identity: a person is nothing over and above a series of more or less connected mental events. He thus denies that there is an especially deep and important sense in which I am the same person as Ivar Labukt of 2030. What gives me a particular reason to be concerned with Ivar Labukt of 2030 is not simply that “he is me”, but certain psychological relations that obtain between us. First, there are direct psychological connections: I may, for instance perform an action then that I plan now, or I may remember something in 2030 that happened in 2010. Also, there will be what Parfit calls psychological continuity: shorter, but overlapping psychological connections between intermediate Ivar Labukts. The relation of connectedness will typically hold to a lesser degree as time passes. One should therefore, he thinks, to some extent discount one’s future interests (*ibid.*, 313-5).⁸⁰

This kind of discount rate must, as Parfit points out, be distinguished from a discount rate with respect to time itself. The interests are not less important because they are in the future, but because a certain psychological relation between me now and me in 2030 will be relatively weak. If the relation hadn’t weakened, future interests would be just as important as present interests. A discount rate with respect to the mere passage of time seems difficult to justify (see section 8.2), and I see no reason to suppose that rational egoism entails or presupposes it.

It is not obvious that the relations of connectedness and continuity have the ability to ground ultimate reasons. Another response to the reductionist picture of persons is to deny the authority of egoistic concern about the future: if there is no deep sense in which persons persist through time, then there is no reason why I should be more concerned with what happens to Ivar Labukt than to any other person. Parfit admits that he cannot demonstrate that this line of reasoning is mistaken (*ibid.*, 307-12). Since, according to rational egoism, I should

⁸⁰ David Lewis defends a similar view, but he thinks it is compatible with the idea that personal identity does matter, since he defines personhood itself in terms of psychological connectedness and continuity (D. Lewis 1976).

not be concerned at all with what happens to other persons, we would then be left with a kind of nihilism about ultimate reasons. At least this is true on the hedonistic interpretation of rational egoism. My hedonic level at this moment would matter to me at this moment, but there would be no point in doing anything about future hedonic levels, since by the time the effects occur I would no longer be the same person.

I will not attempt to decide between these outlooks, since doing so would require a thesis of its own. The plausibility of my arguments for rational egoism in chapters 7-9 is independent of which one of them is correct. What I try to show in these chapters is that it does matter what happens to me right now and that it does not matter (except possibly indirectly) what happens to other people. It is left open in what sense the object of egoistic concern persists into the future. If there are good reasons to believe that it does not persist at all, I would be forced to accept the kind of nihilism outlined above. I guess most philosophers would deny that there are such reasons. I hope they are right.

When it comes to the account of the practical implications of rational egoism given in this chapter, it does matter which of the outlooks we adopt. There is no point in spending time discussing the last one, since it has no practical implications whatsoever. For the sake of simplicity, I will also disregard the possibility of a discount rate. In other words, I assume that rational egoism implies equal concern for all parts of one's life. Any readers who think there should be a discount rate can make the necessary adjustments themselves. I should perhaps point out, though, that a discount rate need not have as dramatic consequences as one might first imagine. As I explain below, a rational egoist would normally develop fairly deep and persistent concerns for the well-being of other people, even if there is no direct rational basis for doing so. If this is possible, one should certainly be able to develop a similar concern for one's future self. For instance, I might refrain from smoking because knowing that I might be giving myself cancer in old age bothers me now, and not (only) because the cancer would bother me then. Similarly, I might work hard on my thesis not (only) because I will enjoy a successful career when I have one, but because I now enjoy looking forward to it.

6.2 The commitment to morality

It is impossible to discuss the practical implications of rational egoism as such. What it would be like to lead the life that best promotes one's interests depends on what one's interests are. I am, obviously, most interested in the hedonistic version of rational egoism. In this section, I discuss in what sense and to what extent it is in one's hedonic self-interest to engage in moral behavior. In the next section, I give an account of the implications hedonistic egoism has for

friendship. Since most theories of self-interest are less individualistic than hedonism, these investigations will yield a sort of minimum egoist commitment to morality and friendship.

In order to assess how an egoist would relate to morality, we must of course be clear about what morality is. As we have seen, some philosophers think that moral principles are just impartially formulated principles for evaluation of conduct, with no necessary claim to independent authority. Others think that such principles cannot properly be called *moral* principles. However, as explained above, they do not thereby deny that there are impartial standards for evaluation of conduct. It is the commitment to these standards, regardless of whether they are called moral, “moral” or impartial, I am interested in discussing here. For the sake of simplicity, I will refer to them as moral principles.

I start by describing the various kinds of egoistic reasons for conforming to moral standards. I will try to show that these reasons are quite strong, and that it is unlikely that hedonistic egoism would, in general, recommend a weaker commitment to morality than the one displayed by actual people in contemporary Western societies. I end the section with a discussion of whom an egoist ought to be treating morally rightly.

Why would it be in one’s interest to perform actions that are morally right? Most obviously, perhaps, there are the directly instrumental reasons: following moral rules is often necessary in order to obtain non-moral benefits or avoid punishment. Even someone who is constantly motivated by self-interest and constantly deliberating egoistically will, therefore, often obey moral rules. This is the kind of justification of the authority of morality Hobbes offers to the “fool” (Hobbes 1651, 101-3). D. A. Lloyd Thomas argues that it is sufficient to solve the age-old question of why one should be moral (Lloyd Thomas 1970; it should be mentioned that he has a very moderate conception of morality). Most philosophers would be fairly disappointed with this solution, however. There is a large number of cases where the immoral agent either is too powerful to fear punishment or too clever to get caught. Also, this kind of commitment to morality is entirely without psychological depth: moral principles figure only as externally given factors in one’s utility calculations.

Moral principles may also be useful rules of thumb or the basis for useful habits. Always having to calculate whether you ought to, say, pay for your groceries would be very tiresome, and you might be tempted to take undesirable risks. It is better to simply pay without giving the matter any thought. These “moral” dispositions would still lack any substantial psychological authority; they would be just like dispositions to brush one’s teeth every day or

to refrain from drinking coffee in the evening. In unusual circumstances, they could easily be set aside.

These directly instrumental egoistic reasons for fulfilling one's moral duty are quite weighty, at least in reasonably well-ordered societies. However, as we will see, there is little doubt that most people will benefit from a more robust moral commitment. I will begin by discussing the kind of moral commitment that is not directly under rational control. Let us call this a psychological commitment. It seems to have two basic components: the general capacity for sympathy and the capacity for a conscience that prescribes or forbids specific kinds of conduct.⁸¹

Most people have a deep tendency to feel good when people around them are happy and bad when they are not. This tendency is not a result of a philosophical belief about ultimate reasons. It is found in very small children, so perhaps it is not even learned (though it may of course be reinforced by praise and rewards). One may possess the capacity even if one is, like the present author, entirely convinced that it has no rational basis.

Because of sympathy, many actions required or encouraged by morality are actually quite pleasant. In particular, this is true for assisting or co-operating with others. (Think of helping an old, grateful lady across the street, for instance.) I suspect that we often underestimate how enjoyable these actions are. Perhaps we tend to assume that, just because someone else benefits from our actions, there must be some loss or sacrifice on our part. The distribution of pleasure is not a zero-sum game, however. I also suspect that the belief in the independent authority of morality may actually lead us to take less pleasure in helping and co-operating than we otherwise would have done. Experiments in social psychology show that people who are given incentives to perform pleasant activities rate the activities as less pleasant than those who engage in them simply for the fun (see e.g. Ariely 2008, ch. 4). Thus, it does not seem too far-fetched to suppose that accepting rational egoism could lead us to enjoy at least some types of pro-social behavior more than a believer in morality with independent authority would.

Another consequence of the capacity for sympathy is that many wrong actions are not particularly pleasant. We do not generally enjoy hurting others. The traditional outlook on morality may encourage the suspicion that if it hadn't been for authoritative moral principles,

⁸¹ This is admittedly an over-simplification: there is presumably no neat and clear psychological distinction between the two capacities, and I do not want to deny that other non-cognitive psychological mechanisms are involved in causing moral behavior. However, the simplification seems safe enough for my present purposes.

there would be nothing stopping us from committing atrocities all the time. (This is reminiscent of how some religious people fear the horrible consequences of atheism.) I don't think there are many people who would derive substantial hedonic benefits from raping, maiming and killing other human beings. It seems to me that the most extreme and disturbing evil actions, at least in modern history, have much more often been motivated by false moral beliefs than by a desire for enjoyment.

Although one cannot, at a given moment, decide how much sympathy to feel, one may of course influence the workings of the capacity over time. Perhaps it might seem as if it would be worth the trouble to diminish or even, if possible, eradicate the tendency to feel good or bad on account of how other people are doing. After all, having this tendency does prevent one from performing certain actions that might have very good hedonic consequences. Also, we should not forget that there is such a thing as sympathetic pain, and getting rid of that would of course be a significant benefit – especially if one is surrounded by a lot of misery. However, it must be kept in mind that a capacity for sympathy has far-reaching hedonic consequences. It is not primarily valuable because of the rushes of sympathetic pleasure we experience now and then. Its general impact on interpersonal relations is much more important. Sympathy allows us to conceive of other people as friends and co-operators, rather than just annoying obstacles to the pursuit of one's own interests. Being “in unity with our fellow creatures”, as Mill calls it, satisfies a deep human need (Mill 1871, ch. 3). Also, it makes others treat us better. In sum, one could say that it raises the baseline hedonic level of social interaction.

Since the quality of one's social life has a huge impact on one's hedonic well-being, it would take some weighty reasons indeed to justify the egoistic desirability of having little or no capacity for sympathy at all. This is not to deny that there are people who feel too much sympathy for their own good, or that there are extreme circumstances where it would be preferable, as judged by egoist standards, to be completely oblivious to the fate of other human beings. However, I do want to deny that it is generally true that people will become happier by blunting their sympathetic sensibilities. It seems much more likely that rational egoism would recommend a change in the opposite direction (at least in the Western part of the world).⁸²

⁸² Sidgwick reached the same conclusion: “enlightened self-interest would direct most men to foster and develop their sympathetic susceptibilities to a greater extent than is now commonly attained. [...] It seems scarcely extravagant to say that, amid all the profuse waste of the means of happiness which men commit, there is no imprudence more flagrant than that of Selfishness in the ordinary sense of the term,— that excessive

More specific dispositions to perform or refrain from performing certain classes of actions, such as theft, promise-keeping or killing, constitute another important source of moral behavior. These dispositions are not just habits: they have a much stronger hold over us. We do not, for instance, keep promises simply because this is what we have always done, but also because we feel in some way compelled to do it, or because the idea of breaking them fills us with disgust or anxiety. I will not discuss here exactly how the human conscience works. The important point for my present purposes is, once more, that it is not under direct rational control; coming to believe that it fails to track ultimate reasons will not automatically silence it. This means that a rational egoist also has access to its benefits. As Sidgwick expresses it, there is a distinction between

the general impulse to do what we believe is reasonable, and special sentiments of liking or aversion for special kinds of conduct, independent of their reasonableness. [...]there is every reason to believe that most men, however firmly they might adopt the principles of Egoistic Hedonism, would still feel sentiments prompting to the performance for social duty, as commonly recognized in their society, independently of any conclusion that the actions prompted by such sentiments were reasonable [...]. (Sidgwick 1907, 173)

This kind of commitment to morality is similar to a phobia. A person standing at the edge of a cliff may experience intense fear without believing that he is particularly likely to fall, in the same way that I still feel bad for breaking my promises even though I am completely convinced that I have no non-derivative reason to keep them. Presumably the moral commitment, too, can be extinguished through cognitive-behavioral therapy. There are some pretty good egoistic reasons for not doing so.⁸³ First, a conscience keeps its owner out of a lot of trouble. Many immoral actions promise instant gratification, and habits and rules of thumbs are often insufficient to prevent us from taking undesirable risks or imprudently giving in to temptation. Psychopaths, who lack these firm moral dispositions but are perfectly capable of calculation, frequently end up in jail, broke or dead. Second, I think that having a deontological conscience also provides general hedonic benefits in social interaction. It partly constitutes and partly encourages a sentiment of *respect* towards other human beings. Respecting a person is not the same as caring for his well-being; it is more like conceiving

concentration of attention on the individual's own happiness which renders it impossible for him to feel any strong interest in the pleasures and pains of others." (Sidgwick 1907, 501)

⁸³ Utilitarians also think it is desirable to have aversions against certain kinds of conduct, such as against lying, even if there is nothing intrinsically immoral about lying according to the principle of utility (see e.g. Sidgwick 1907, book IV; Hare 1981, ch. 2-3).

him as an independent source of claims or side constraints on what you may do. This sentiment makes interpersonal relationships deeper and ultimately more rewarding.

I suspect that it is more common that human beings have too much commitment to specific moral rules than too much sympathy, as seen from the point of view of hedonistic egoism. For instance, some people are too concerned with telling the truth for their own good. Perhaps hedonistic egoism would, in general, recommend a somewhat looser psychological commitment to specific moral injunctions. I don't think the changes would be very significant, though.⁸⁴

My claim that hedonistic egoism would have relatively conservative implications for our psychological commitment to morality is strengthened by the fact that we do not get to design our own psychological commitment from scratch. Part of it is presumably attributable to our genetic make-up. It is heavily shaped by parents, teachers, peers and the general culture during childhood and adolescence. This means that our character is formed more in accordance with moral standards than egoistic.⁸⁵ (Even if the people around you should be rational egoists, this does not mean that they will provide you with the set of dispositions and sentiments that best promotes your interests. Rather, they would encourage the psychological commitment to morality that is in *their* best interest, which is likely to be more substantial.) By the time one is mature enough to reflect on one's own dispositions, they are so deeply engrained that it will take quite a lot of pain and effort to change them. For this reason, it may be a good idea not to try to correct one's psychological moral commitment, even if it should be somewhat stronger than what one ideally would want.

Most people also display what we may call a deliberative commitment to morality: they reason about what to do and make explicit moral choices. This commitment goes beyond the egoistic psychological commitment in two ways. First, it typically aims for the actions that are in some sense really right, and not just considered to be right by one's parents or popular opinion. Our sympathy and conscience, on the other hand, can serve their egoistic purpose

⁸⁴ Jan Österberg also points out that the egoist must "use a set of derived normative rules to guide his daily life" and thinks that "these rules – which, to some extent, must be internalized – largely coincide with common-sense morality" (Österberg 1988, 103).

⁸⁵ According to R. M. Hare, there is actually no difference between the moral dispositions you would want to instill in children in order to ensure that they are likely to have a happy life, and the dispositions that the principle of utility would recommend (Hare 1981, 194-8). This is surely an exaggeration, but the difference may not be very great.

without tracking any kind of moral truth. A Nazi may get pleasure from being in unity with his fellow Nazis, and in certain societies it will definitely be a good thing to have a conscience that demands immoral actions like honor killing and forbids morally innocuous things like dancing or homosexual relationships. Second, if you have a deliberative commitment to morality, your moral beliefs will sometimes play some role in bringing about morally right actions. If you, say, give money to charity, then you will to a certain extent be moved by the belief that doing so is morally desirable, and not only by a sympathetic impulse, a bad conscience or a desire to develop an egoistically useful “helping disposition”.

Having a deliberative commitment to morality provides egoistic benefits. First, other people will tend to like you better. (It is of course true that you could achieve the same effect just by pretending to have the commitment, but doing so would be tiresome and risky.) Second, doing what one, on a cognitive level, takes to be morally right is an attractive non-egoistic project. Just as we may bring about more pleasure by aiming for other things than pleasure, we may bring about more of our own pleasure by being engaged in promoting the hedonic well-being of others. The moral project is a comprehensive and lasting way of providing structure to our lives. Also, it is easily shared with other people. Both these features contribute to its egoistic value. In addition, there is something particularly appealing about the content of this project. We like the idea that we could, during our short stay here, make the world a slightly better place (see e.g. Singer 1993, 332-5). Or, as Gauthier puts it, we enjoy cooperating with others on fair terms, even if our ultimate goal may be non-cooperative in nature (Gauthier 1986, 330-9). This is not just because others will show their gratitude or reciprocate; they may not always do so. The mere knowledge that we have treated other people in ways that they could not reasonably reject is a source of satisfaction independently of what these other people actually say and do.⁸⁶

In section 5.3, we saw that rational egoism does not conflict with the universal validity (as opposed to the independent authority) of moral principles. Thus, we need not worry that a rational egoist will lose interest in having a deliberative commitment to morality because he thinks that his moral concerns are ultimately just arbitrary reflections of his own psychological make-up or social environment. However, one might perhaps think that one needs to believe that moral principles have independent authority, and not just universal content, in order to derive any substantial satisfaction from acting on them. This is not the

⁸⁶ Thomas Scanlon puts great emphasis on this value in his account of the authority of morality, though he does not construe it in hedonic terms (Scanlon 1998, ch. 4).

case. As explained in section 4.2, human beings have the ability to engage in and enjoy many different pursuits even if they do not think that the pursuits in question are based on or constitute sources of ultimate reasons, and even if they are not engaged in any deep form of self-deception. I see no reason to suppose that morality should not be able to attract us in this way. As Butler points out, there is nothing particularly problematic about an egoist's commitment to morality in this respect: it should be no more surprising than a commitment to seeking power, wealth, fame or any of the things that we easily imagine egoists being concerned with (Butler 1725, 30-6).⁸⁷

In fact, there is reason to believe that an egoist's deliberative commitment to morality would tend to take priority over deliberative non-egoistic projects that do not directly involve the interests of other people, such as, say, being a good philosopher. At least this is true on a hedonistic interpretation of egoism. As a hedonistic egoist I would know that, from a philosophical point of view, being a good philosopher does not, in any way, have intrinsic value. However, taking into account the interest of others is not meaningless in the same strong sense, since what I do really matters *to them*. This makes the project of treating them rightly seem much more attractive. If I became convinced that the people around me were zombies lacking phenomenal consciousness, I would find it harder to derive satisfaction from treating them well.⁸⁸

Those who believe that morality is overriding, or at least has considerable independent authority, will of course display a stronger deliberative commitment to morality than believers in rational egoism. They would presumably also derive deeper and more permanent satisfaction from their commitment. However, they would also be prone to making substantial long-term sacrifices of their own happiness. Perhaps some people are lucky enough to find themselves in circumstances where the increased satisfaction would outweigh the required sacrifices. These people would have most reason (not in the epistemic sense, of course) to be

⁸⁷ C. H. Whiteley acknowledges that an egoist could pursue various projects besides his own happiness, but claims that morality is not one of them: it cannot be "a self-subsistent, independent interest like his interests in chrysanthemums or golf or the love of a good woman" (Whiteley 1976, 96). He never explains why the commitment to morality must be more superficial than an interest in pretty flowers.

⁸⁸ Incidentally, it is possible that accepting rational egoism will diminish the attraction certain people feel towards deeply immoral acts. Perhaps some of these people are motivated in part by an idea that being evil makes one very special. If they had been rational egoists, they would have known that from a philosophical point of view there is nothing special or heroic about rejecting moral principles in order to be, say, a serial killer. The only thing that distinguishes a serial killer from the rest of us is that he derives his pleasures from an unusual source.

supporters of some version of moral rationalism. However, I strongly suspect that for most of us, the kind of deliberative commitment to morality that is compatible with a belief in rational egoism will be preferable. Also, it should be noted that if you already are convinced that rational egoism is true, it will be very difficult to get rid of this conviction in a psychologically healthy way, and so you might have to stick to it even if you would have been better off if you had not become a rational egoist in the first place.

If my claims so far in this section are at least roughly right, a supporter of rational egoism would in many situations be difficult to distinguish psychologically from a person who thinks that morality has independent authority. The egoist will feel impulses to help others and to engage in or abstain from certain more specific kinds of conduct. These impulses often have the character of what Scheffler calls “authoritative motivation” (Scheffler 1992, ch 5): one’s very perception of a situation is motivationally laden. Moral actions come to be represented as actions that just have to be done – and this happens without explicit reasoning or any general desires to benefit others or to do what is right.⁸⁹ On a more cognitive level, the conviction that an action would be morally right automatically counts as a weighty consideration in its favor; in everyday deliberation, moral reasons do not have to be constantly ratified by egoistic calculation.

This similarity to a believer in moral rationalism will, however, disappear when the situation is unusual and when a lot of the agent’s pleasure is at stake. It is not a good egoistic policy to make major decisions directly based on moral principles. This is not to deny that moral concerns may play an indirect role: the hedonic consequences of a commitment to morality must of course be taken into account in the egoistic calculation. It is sometimes hard to be moral, but in a large number of cases the egoist may conclude that the price is worth paying, since there will be long-term benefits of the various kinds discussed above. Some people may even find that they have most reason to start working for Oxfam because of these benefits. My present point is just that it would be irresponsible to do so simply on moral impulse.

⁸⁹ Scheffler thinks that the authority in question is not only psychological, but also normative: there is a genuine reason to do what our conscience (or *superego*, as he calls it) prompts us to do in this fashion. However, he never provides any good argument for why what seems to be just a psychological mechanism is a source of ultimate reasons, nor does he properly examine the normative consequences of this view. (Are neurotic people just people with particularly strong moral reasons?) For discussion, see Brink 1994; Darwall 1995; Scheffler 1995.

I suppose relatively few will find it in their interest to work for Oxfam, or to engage in anything else we may reasonably refer to as heroic moral action. However, at the risk of sounding repetitive, I do not think that the level of deliberative commitment to morality recommended by rational egoism would be very different from the one displayed by most actual people.⁹⁰

Some authors seem to think that the fairly moderate, on-and-off attachment to moral principles suggested here is not feasible: in order to derive maximum enjoyment from being moral, one must have a stronger deliberative commitment. Gregory S. Kavka, for instance, apparently holds that an egoist has to choose between being completely amoral and adopting “the moral way of life” to such an extent that he would be willing to sacrifice his life in order to fulfill his duties (Kavka 1985, sect. II-III). Now a believer in rational egoism could no doubt sacrifice his life as a result of strong impulses or emotions, and for some people in some situations death would be preferable to saving one’s life in an immoral way. However, Kavka seems to think that the egoist, if he is to have access to genuine moral satisfaction, would have to be willing to consciously decide to sacrifice his life even if he could go on living a life that contains more happiness than unhappiness. I cannot see why the egoist’s commitment to morality needs to be this strong. We are clearly able to derive enjoyment from other projects (say, having a good career) without letting the projects trump any opposing consideration. The behavior of millions of actual people seems to show that we can do the same when it comes to morality.

Neither can I see how the egoist’s commitment to morality *could* be this strong. Sidgwick is, it seems to me, right in holding that when morality demands extreme sacrifices, any “sane person, who still regards his own interest as the reasonable ultimate end of his actions” must be able to “deliberate afresh, and to act (as far as the control of his will extends) without reference to his past actions.” (Sidgwick 1907, 174) A believer in rational egoism could not consciously decide that he ought to give up a life worth living (though he may end up doing so as a result of weakness of will, of course).

David Gauthier’s constrained maximizer has a similarly rigid commitment to moral fairness. He makes “a choice about how to make further choices; he chooses, on utility-maximizing grounds, not to make further choices on those grounds” (Gauthier 1986, 158). A

⁹⁰ Of course it is one question how one ought to relate to morality according to rational egoism and another question how actual people would react if they became convinced that the theory is correct. Presumably many would overlook the considerations in favor of being a moral person I have presented here. In *that* sense, rational egoism would probably have a subverting effect on morality.

constrained maximizer will sometimes, if he misjudges other agents' willingness to cooperate, find himself in situations where his disposition yields suboptimal results (ibid., 169). It is important to stress that he will not just be making short-term sacrifices that will be compensated by long-term gains such as attaining a desirable character or being trusted by others. The constrained maximizer will be knowingly choosing what is, on the whole, worse for him. This seems like an unnecessary strong moral commitment. Also, it is completely psychologically unrealistic. Even if it should be desirable in certain cases involving cooperation, we do not have the ability to force ourselves to make certain choices in the future simply by making a resolution. As Sidgwick says, we can always "deliberate afresh". No matter how much I try to constrain myself today, I will not be psychologically compelled to give up a net egoistic benefit tomorrow. In order to make sure that I give up the benefit, I would have to deceive myself in the deep and psychologically unhealthy sense.⁹¹

I will consider one more question about the egoistic commitment to morality, and that is whom it includes. There are actually two different questions here:

(1) Against whom does one have moral duties?

(2) Whom does one have reason to treat morally rightly?

⁹¹ Richard Joyce defends an error theory about morality (Joyce 2001, ch. 1-5), but thinks we would be wise to keep the idea of a morality with overriding authority as a "fiction" in everyday life (ch. 7-8). In many ways, his suggestions are similar to mine. He does, however, think that most people would find it in their interest to stick to the fiction of an overriding morality in almost all situations, perhaps with the philosophy classroom as the only exception (ibid., 221). Again, it is difficult to imagine someone actually living his life in this way. Could a sane person "forget" his deepest normative beliefs when making important and difficult choices such as whether to, say, have an abortion or go to work for charity? I certainly would not be able to do this. Also, it seems very risky to have a disposition of this kind. How do I know that I will not end up sacrificing my interests or even my life for the sake of morality? Joyce may be right that strong fictionalism would be better than constant egoistic calculation or reliance on blind habit (ibid., 8.3), but it does not seem more attractive than the weaker commitment to morality outlined here. To be fair to Joyce, he does point out that his estimate may be mistaken and that fictionalism in a weaker form is also a possibility (ibid., 221). Also, it should be noted that Joyce is not a rational egoist, but an instrumentalist about reasons, and so his justification for adopting a fictionalist stance towards morality is that doing so is the best way of satisfying our preferences (ibid., 177). Perhaps many people have sufficiently strong altruistic preferences for a strong form of fictionalism to be the best option. (Joyce does not rely on this claim in his argument.)

The answer to (1) is determined by the correct theory of morality. According to utilitarianism we have, in principle, equally strong duties towards all creatures capable of suffering or happiness. Other theories are more selective; they hold that our duties to certain people are stronger than others, and perhaps also that there are people who have no moral claim on us at all (because, say, they have violated our rights). I do not want to take sides in this debate; rational egoism does not have any direct implications for this issue.

(2) is a different question, though. Even if I have equal moral duties towards all human beings, it may turn out that it is only in my interest to be “in unity” with some of them. Let us consider some possible limitations on the scope of the moral project.

Most actual people show greater concern for people who are close to them (what matters is social distance and not physical, of course). On the psychological level, a moderate limitation of this kind seems egoistically desirable. Feelings of guilt and sympathy are more easily aroused when dealing with people who are, in some respect, similar to oneself, and one’s moral efforts are more likely to be reciprocated. If one lives in a society where different classes or races are clearly segregated, it might even be in one’s interest to reduce one’s psychological moral commitment to members of other social groups as much as possible. (This condition is, of course, not satisfied in contemporary Western democracies.)

Given that one has no false beliefs about the relative moral worth of people, this kind of limitation may seem to be excluded in the cognitive commitment to morality. It does seem very peculiar to make a deliberate decision to be strongly concerned with the moral rights of one’s own social group and completely disregard the moral rights of anyone who happens to belong to a different group. However, a more moderate favoritism does appear both feasible and reasonable. For instance, a person who has been abused as a child may want to focus his moral efforts on helping people in a similar situation, even if he does not believe that the contribution he can make to this cause is more important from a moral point of view than the contribution he could make to other worthy causes, such as the fight against poverty.

People can be distant in a temporal as well as social sense. Would an egoist be interested in fulfilling duties to future generations? Again, it is not clear how extensive the duties in question are. In any case, however, I think it is reasonable to suppose that most egoists would have a fairly strong wish for the human race to survive and prosper. Some of this altruism would be very partial; one would be concerned about the welfare of one’s descendants. However, I also think that the idea of making some contribution to (or at least not obstructing) the “project of humanity” is an attractive non-egoistic project. There is, of course, a limit to how committed one should be to this project (I guess hardly anyone would find it in their

interest to give up basic goods), but I do believe that most people would have quite strong egoistic reasons for making the fairly moderate sacrifices necessary for, say, controlling global warming and reducing global poverty.⁹²

How would an egoist treat animals? If a hedonist axiology is accepted, there would presumably be no important moral distinction between human beings and animals.⁹³ Now I do think that rational egoism generally recommends having a moral concern for animals. (Partly this is an unavoidable effect of the attitude towards human beings it recommends; if you are able to bond with people and feel sympathy for them, you will tend to do the same when it comes to animals.) However, I see no reason for supposing that this concern should always be as strong as the moral concern for human beings. For most of us, it is easier to isolate ourselves from the suffering of animals than the suffering of other people, and animals (except for dogs, perhaps) do not have the ability to reciprocate our concern for them to any great extent.

Peter Singer has labeled our cavalier attitude towards animals speciesism, and thinks it is morally objectionable (see e.g. Singer 1993, ch. 3). That may well be true. But as we have seen, Singer also seems to accept rational egoism. And a rational egoist needn't justify his relative lack of concern for animals with a claim that human beings have supreme moral worth; he could simply point out that he gets more happiness from being in unity with his fellow human beings than from being nice to cows and chickens.

I will consider one final candidate group to be excluded from the egoist's moral concern: immoral people. It may not be morally right to disregard the interests of people who treat others badly. (According to classical utilitarianism, for instance, it isn't.) Nevertheless one may not find it in one's interest to care about treating such people rightly. The appeal of treating others in ways that they could not reasonably reject is to a large extent dependent on the supposition that they would do the same for us. I think that if people were either good or bad, and if the bad ones looked and acted very differently from the good ones, an egoist belonging to the good kind would not want to develop any significant moral concern for bad people. He should simply disregard their interests.

⁹² Or more precisely: they should be willing to do so as long as other people do it as well. The solution to these problems would have to be political. This is hardly an unusual view; again it seems to me that most actual people behave in a way that is roughly compatible with rational egoism.

⁹³ Though it is not entirely clear how certain we can be that animals have experiences with hedonic tone (see Allen, Fuchs, Shriver & Wilson 2005; Panksepp 2005).

Of course, in the real world these conditions aren't fulfilled. First, it is notoriously difficult to determine the moral worth of people, and those who have actually attempted to weed out the immoral ones have all failed in this respect. (They thought that the bad people were the Jews, infidels or political dissidents.) Second, even if it should be possible to determine exactly who the bad guys are, it would presumably turn out that they look just like other people, and it would be very difficult to maintain a psychological moral commitment towards only the good ones. There is, then, no easy way of adjusting one's moral concern to the moral worth of people. Nevertheless I do think that rational egoists would be able to strike some kind of balance between indiscriminate niceness and completely disregarding the interests of people believed to be immoral.

It should be noted that the egoistic reasons for being (reasonably) moral I have given in this section do not apply to all agents. First, certain individuals are too powerful to have any directly instrumental reasons to be moral. There may perhaps also be situations where a psychological commitment to morality is egoistically undesirable. More importantly, there are people, such as psychopaths, who seem to lack the capacity to develop this kind of commitment. This is, generally speaking, bad for them, but given that they do lack the capacity, they have no reason to try to be more moral. Something similar applies to the deliberative commitment to moral principles: it appeals to most of us, but not all. If it does not appeal to you, you are not failing to recognize the reasons for action you really have. What happens, rather, is that your constitution prevents you from having the reasons in the first place.

It may be thought that this limitation undermines my claim that rational egoism implies that we should not be significantly less moral than we are. I don't think this is true, since I suspect that, as things are, those people who are not favorably inclined towards morality tend to behave just as badly as rational egoism would tell them to. It is at best, I believe, extremely rare that people who have no conscience or natural feelings of sympathy, and who are in no way immediately attracted to the idea of making the world a better place, force themselves to lead a morally good life because of a philosophical conviction that moral reasons have independent authority.

We are now perhaps in a better position to decide whether principles without independent authority may count as genuinely moral principles. Some of the resistance to this idea may be

attributable to the worry that one could only have a very superficial commitment to principles of this kind. This is, as we have seen, not true.

David Brink thinks that the fact that a rational egoist could have much the same motivational relation to morality as a moral rationalist suffices to meet the “Kantian” objection that complying with morality on grounds of self-interest would be heteronomy (Brink 1989, 244). However, what Kant requires from a genuinely moral person is not that some particular drive or feeling must be present whenever he acts, but that he acts on the basis of a conviction that morality has supreme authority (see Wood 2008, 25-6). And this a rational egoist cannot do, of course. Even though he might acknowledge that the needs fulfilled by morality are deep and widely shared, they are not in principle different from other needs. There is no special authority that attaches to morality and not to, say, sports or fine food.

It should be clear, then, that the commitment to morality I have been describing in this section is not all that Kant and many others have taken the moral commitment to be.⁹⁴ On the other hand, it may be more than what some have taken a rational egoist’s commitment to morality to be. It is simplistic and misleading to say that a rational egoist cannot be a moral person since he only cares about himself. I do not find it necessary to discuss exactly what we should say. What matters is what the rational egoist’s commitment to impartial principles is like and not which words we use to describe it.

6.3 The commitment to friendship

The moral commitment discussed in the previous section does not capture all the concern human beings show for each other. We also have more specific relationships to specific people. The only case I will discuss in any detail is friendship between equal adults, but I take it that similar considerations apply to other particular relations, such as those between spouses or relatives.

As we will see, it is probably true that a person who believes that friendships ground ultimate reasons would derive somewhat deeper satisfaction from his friendships than a hedonistic egoist. Perhaps there are cases where this belief can be held without any substantial hedonic losses. In these cases, hedonistic egoism entails that there is most reason (in the sense relevant for action) to believe that friendship grounds ultimate reasons, and so a person acting

⁹⁴ Peter Singer makes the same concession in his discussion of the egoistic commitment to morality (Singer 1993, 322-6).

in accordance with hedonistic egoism will obviously be capable of genuine friendship. However, the interesting question is of course whether a person who *believes* in hedonistic egoism can be a real friend.

Unfortunately there has been a tendency in the literature to treat this as a question that should be answered with a simple “yes” or “no”. In this section, I will argue that a believer in rational egoism can fulfill many of the requirements that could reasonably be said to be implicit in the notion of friendship, but not all of them.

Some writers argue that one ought, according to egoism, only to develop friendships of a directly instrumental character. John van Ingen thinks that an egoist must consider other people as mere “obstacles” or, in case he can benefit from them, “resources” (van Ingen 1994, 107-8). This directly instrumental attitude is, of course, incompatible with any deep form of friendship (ibid., 171-2). The conclusion van Ingen draws from this is that we should not be egoists (ibid., ch. 8). Max Stirner had a similar view of the nature of egoist friendship, but he seems perfectly willing to accept it: for him, one’s friends are just “useful bodies” to whom one should not in any way feel obligated (Stirner 1845, 266, 287-97).

I do not want to deny that this approach to friendship may be egoistically desirable for people who find themselves in extreme situations or lack the psychological capacity to bond with others. However, for most people in most circumstances, it clearly is not. Deep friendships constitute a very long-lasting and valuable source of pleasure (and protection from pain), and though there is of course a hedonic downside (one will sometimes miss one’s friends, feel bad for not helping them, feel distressed when they are in pain etc.), the benefits normally outweigh the costs.

The fact that deeper forms of friendship are egoistically desirable does not mean that a believer in egoism will be capable of attaining them. Sometimes it is claimed that a rational egoist cannot be a good friend because of the way he *behaves*. William H. Wilcox writes that “the problem with the egoistic hedonist’s overriding commitment to his own pleasure is that the practical effects of this commitment are so pervasive that little room is left for a concern for another’s well-being to have much practical effect.” (Wilcox 1987, 78-9) This is an implausible claim. As pointed out before, hedonistic egoism is not intended as a decision procedure to be followed in everyday life. A sensible egoist is thus likely to meet Wilcox’ requirement that “one’s concern for a friend *usually* have a conclusive effect on one’s practical deliberations affecting the friend’s welfare.” (ibid., 78)

It should not be denied that it sometimes would be egoistically reasonable to engage in explicit egoistic calculation in order to decide how to treat one's friend. However, this does not mean that the resulting action will fail to benefit the friend. There are weighty egoistic reasons for being nice to one's friends. Most obviously, perhaps, it is usually quite pleasant. For those of us who are not self-centered, most of the small "sacrifices" we make for our friends (cooking them dinner, picking them up at the airport etc.) aren't really sacrifices at all; we generally enjoy being engaged in these activities. Of course it is now and then genuinely painful to help a friend. But we must remember that refraining from doing so would typically also have negative effects. First, it would tend to weaken the friendship and so deprive us of future satisfaction. Second, one would presumably feel bad for abandoning a friend in need. Even if one decided to break off the relationship, such sentiments would typically remain for quite some time. It is true that according to rational egoism there would be no philosophical justification for these feelings, but that wouldn't make them any less real. Third, how one treats a given friend also has some influence on one's more general way of relating to other people. It is very difficult to behave cynically selfish towards certain of our friends and at the same time have genuine and rewarding friendships with others. Being a good friend is to a certain extent a way of life, and not just a policy adopted towards a specific set of people.⁹⁵ Finally, believers in rational egoism will tend to behave more friendly than they should through weakness of will. It can be very hard to force oneself to perform actions that hurt those one cares about, especially if the long-term benefit for oneself is only minimal.

I do not mean to suggest that a rational egoist will be the most devoted friend imaginable. My claim is only that his, given the behavioral requirements actual people make on their friends, his conduct would generally be found satisfactory. If there is a serious incompatibility between rational egoism and friendship, this will have to be because of what goes on in the head of the egoist, rather than because of what he does. One worry might be that the egoist will have an unacceptable motivation for his friendly behavior. However, as is hopefully clear by now, it would be a mistake to assume that a believer in rational egoism is also a person

⁹⁵ Elinor Mason thinks that a sophisticated impartial consequentialist (the point presumably also applies to the sophisticated egoist) would never calculate the utility of a single relationship, but only the utility of having a completely general "pro-friendship attitude" (Mason 1998, 389-90). This seems to me exaggerated; we do not have to choose between never evaluating single relationships and doing so continuously; ending one relationship because it is suboptimal does not automatically prevent you from enjoying others. Nevertheless there is an important truth in the claim that one needs to have a broader perspective than just each single relationship when deciding what to do for a friend.

who is always, or usually, explicitly motivated by self-interest. In everyday friendly interaction, egoistic considerations need not enter his mind at all (see Railton 1984, 141-6).

Another worry is that a belief in egoism might undermine the emotional responses necessary for friendship. Michael Stocker thinks that an egoist cannot, unless he deceives himself, feel real concern and respect for others (Stocker 1976, 456-58). This claim presupposes an overly rationalistic conception of emotions. Deep psychological features such as the capacity to love or care do not go away as a result of an abstract philosophical conviction.

It is primarily on the level of philosophical belief that there is a tension between egoism and friendship. One question is whether a belief in egoism would interfere with the pleasures of friendship. There is no need to dwell on the conviction that rational egoism is true, but it will presumably from time to time pop up in the egoist's mind at inconvenient moments. This problem may not be very serious (judging from my own experience, it isn't). However, what I want to discuss is not how much pleasure an egoist could derive from friendship, but in what sense and to what extent he can actually be said to be a friend.

Even if his fundamental normative belief does not preclude the egoist from treating his friends well or from experiencing the emotions appropriate to a friend, it may still preclude him from being a real friend (see Kapur 1991; Cocking & Oakley 1995, 99-102). In order to see this, imagine that you have a friend who believes in rational egoism. Since he has a well-developed capacity for sympathy and care, and since he understands all the egoistic benefits of friendship, you are not worried that he will ever actually treat you badly or fail to feel what you would like him to feel. However, some of his normative beliefs about your friendship may nevertheless disturb you. Let's say that you present him with the following scenario: "An evil demon offers to slightly increase your hedonic level for the next hour if you let him torture me for a year after you die. If you agree, the demon will make you forget that you ever talked to him." Your friend will have to admit that he thinks he ought to accept such an offer, and this may lead you to wonder if he really is your friend – even if you are perfectly certain that there are no evil demons.

How serious is this problem? This depends on what exactly it is that the hedonistic egoist believes about friendship, and how this belief differs from other beliefs that one could reasonably expect a friend to have.

A supporter of hedonistic egoism believes that his friends are not a direct source of ultimate reasons. He does not believe that this is so because his friends are somehow deficient as human beings. If he did, a hesitation to call him a real friend would be very understandable

indeed. Rather, his doctrine is supposed to apply to all agents, and he would acknowledge that his friends have no more reason to be concerned with his well-being for its own sake than he has for being concerned with theirs. Also, we must not forget that hedonistic egoism only denies that the friendship itself generates ultimate reasons, and not that the enjoyment one experiences when being friendly is intrinsically valuable. There is no reason why an egoist should conceive of friendships as just a means to some clearly non-friendship-related good, such as money or social status.

What does a true friend believe about friendship? According to G. E. Moore, he ought to believe that friendship is an intrinsic value, or, more precisely, that pleasure, contemplation of mental qualities in a friend and the existence of the qualities themselves form an organic whole whose value is much greater than the pleasant experiences considered in isolation (G. E. Moore 1903, 237-8, 251-3). If this is correct, then friendship is not necessarily in conflict with egoism. An egoist could hold that being involved in such relationships is part of what is in his interest, and that there is, therefore, an ultimate reason to promote and protect his friendships. However, there would be a conflict with hedonism. A hedonist could not attribute intrinsic importance to friendships, but only to the enjoyment they provide. This is true even if he thinks that the hedonic states of other people than himself are a source of ultimate reasons.

On the other hand, many authors believe in ultimate reasons that do not concern our friendships as such, but rather our friends (see e.g. Kapur 1991, 483-4; Scheffler 1997). We should not (only) promote – or honor or respect or whatever one thinks appropriate – the relationships themselves, but (also) the interests of our friends. Imagine, for instance, that a friend needs some money so he can go to Hollywood and become a movie star. If you lend him the money, he will succeed and the two of you will gradually drift apart, but your friend will nevertheless have, on the whole, a better life. If you don't lend him the money, this will not significantly harm your relationship; you will go on being close friends as long as you live. To the extent that you value the friendship itself, it is clear that you should not help your friend. To the extent that you value your friend's well-being, on the other hand, it is clear that you should.

On the latter understanding of friendship reasons, the hedonistic egoist's potential failure to be a friend derives from egoism and not hedonism. As a hedonist, nothing prevents him from accepting that he has reasons that are neither egoistic nor impartial to be concerned with the pleasures and pains of his friends. As an egoist, something obviously does.

I will not attempt to decide which of these reason claims can most plausibly be said to be implicit in the notion of genuine friendship (they are not mutually exclusive, of course). Whichever one picks, it differs significantly from the claims made by the hedonistic egoist.⁹⁶

To what extent should this make one hesitant to call an egoist a friend? The answer to this question depends on another question which is curiously absent in the debate on the compatibility between egoism and friendship: Is egoism true? If it is not, and there really are ultimate reasons based on friendship, then *of course* we will think that anyone who disregards them must in some important sense be deficient as a friend. Part of the common reluctance to accept the egoist as a real friend surely stems from the conviction that he is somehow *in error*, i.e. that he fails to acknowledge claims from his friends that are *valid*.

On the other hand, if rational egoism is true and friends cannot make claims with independent normative authority on one and other, then it is much harder to see why you should be offended by your friend's egoistic convictions. As long as he treats you well and feels the emotions appropriate to friendship, there is nothing missing that could reasonably be required to be present. The egoist makes no mistake in how he treats you.

Of course one may feel a bit sad that the kind of ultimate reasons-based relationships that would have been possible on a different theory of reasons is not available to rational egoists. However, it seems, given that rational egoism is true, overly dramatic to insist that there can be no genuine friendship between two rational egoists. At this point, it may be worth pointing out that Aristotle, whose distinction between instrumental and genuine or "perfect" friendships (*Nicomachean Ethics*, 140-4) is often cited, may plausibly be interpreted as a rational egoist.⁹⁷ But ultimately our choice of words will depend on the notion of friendship

⁹⁶ There is a closely related debate on whether friendship is compatible with a belief in impartial versions of consequentialism, such as utilitarianism. If Moore is right that friendships have great intrinsic value, then there may not seem to be any conflict between being a friend and being a consequentialist (see Henden 2007). However, an impartial consequentialist would think that he has just as much reason to promote other people's friendship as his own (though there may, of course, be pragmatic reasons for focusing on his own relationships). This is not what we expect friends to believe. In fact, the rational egoist, who would only be concerned with his own friendships, seems to do better in this respect.

The impartial consequentialist is closer to fulfilling the requirement that one should be directly concerned with what happens to one's friends. Although he cannot, in principle, give them higher priority than himself or anyone else, at least he gives them some priority. In this respect, a consequentialist friend thus seems preferable to an egoist friend. (On the other hand, you may perhaps be disturbed by the fact that your friend would continue the relationship even if he didn't enjoy it, as long as he thought it would be sufficiently enjoyable for you.)

⁹⁷ He was of course not a hedonist.

we started out with. The aim of this section has been to describe certain relationships a believer in egoism could have with other people. It is of little importance whether we call these relationships “genuine friendships” or not.

7. The case against ultimate reasons

As explained in the introduction, the basic idea defended in this thesis is that the anti-realists are right about all reasons except egoistic hedonic reasons. In this chapter, I give some general arguments for normative anti-realism. These arguments are not intended to be original; I will, at least for the most part, stick to what is by now considered standard arguments for the position. I will, however, criticize some responses that have been made on behalf of realism. Discussing these issues in the detail that they deserve would presumably require a thesis of its own. In this chapter, my aim is only to show that there at least appears to be some quite weighty reasons for doubting whether there are any ultimate reasons.

7.1 Queerness

J. L. Mackie famously claimed that “objective values” would be very queer and mysterious entities, “utterly different from anything else in the universe” (Mackie 1977, 1.9). In order to assess this objection, it is important to be clear about what an objective value is supposed to be. Mackie is not particularly clear about this himself, and this is reflected in the various responses his argument has provoked.

At one point Mackie claims that objective values have necessary motivating power of all agent who know about their existence: “an objective good would be sought by anyone who was acquainted with it.” (ibid., 40) He even suggests that the motive in question would have to be “overriding” (ibid.). I think Mackie is right that this kind of magnetic force would be quite queer, but it is not a necessary part of the realist picture. Most normative realists are also externalists: they do not think that a judgment that one ought to do X must be accompanied by some form of motivation to do X in order to count as genuine and sincere (see e. g. Brink 1989, ch. 3; Shafer-Landau 2003, ch. 7; Tännsjö 2010, 43-6). Of course, normal human beings are built in such a way that they will in almost all cases be at least somewhat inclined to do what they think they have reason to do, but this is just a fact about our psychology and not a reflection of any form of conceptual or metaphysical necessity.

As pointed out in section 5.3, some moral realists think that morality need not be a source of ultimate reasons. If this is accepted, it may not be very difficult to defend *moral* realism against the queerness objection. The realists in question typically point out that moral facts presumably supervene on natural facts in much the same way as, say, biological or economical facts supervene on physical facts (Boyd 1988; Brink 1989, 7.1). There is, at least

for the moment, no way of reducing statements in biology or economics to statements in physics, but that is not taken to mean that the former disciplines are metaphysically obscure or that they lack universal validity. This analogy between morality and empirical sciences may or may not be convincing. The important point for my purposes is that it would only show that morality could be objective in the sense that moral judgments have universal validity. It would not (as the authors themselves admit) show that morality has independent authority, since no such authority can be found in biology or economics.

It would seem, however, that this kind of authority is an important part of what Mackie has in mind when he talks about objective values (see Joyce 2001, 30). At any rate, I believe that it is the special normative authority of objective values, or ultimate reasons, that first and foremost would be potentially queer (see Garner 1990; Hampton 1998, 45-9). Of course, if a naturalist version of normative realism could be made to work, then this authority would turn out not be particularly queer at all. I will briefly explain why I do not believe this way of responding to the queerness objections will work. The worry is not that naturalism as a metaphysical view could not be correct; I can certainly see why philosophers would be wary of postulating non-natural facts. (I return to this issue in the next chapter.) Rather, the problem is that if naturalism is correct, there couldn't, as far as I can see, be any ultimate reasons.

Naturalism about reasons comes in analytical and non-analytical versions. I will focus on the latter, since it is generally thought to be more promising. The idea, then, is that even if our normative concepts may have a non-natural meaning, the properties that they pick out in the world are nevertheless just natural properties. Now there is a not particularly interesting sense in which pretty much all moral philosophers would agree with this doctrine: they believe that values reside in or that reasons are grounded in natural objects. The things that there is a reason to seek, such as pleasure, accomplishment or friendship, need not be understood in a non-natural way. If moral philosophers are non-naturalists, it is because they think these naturalistically respectable things have a *further* property of being valuable, or something there is a reason to seek, that resists naturalist treatment. As Parfit says, we must distinguish between the properties that make something valuable and the property of being valuable itself (Parfit forthcoming, sect. 87).

Some naturalists seem to overlook this distinction. Allan Gibbard thinks that questions about what one ought to do are questions about one's plans for actual or hypothetical situations. Also, he thinks that such plans can be "right" in what appears to be very close to a realist sense, but he thinks that this is entirely compatible with naturalism. In fact, he seems to think that naturalism *follows* from our need for planning:

normative terms like ‘ought’ refer to [...] properties and relations that are natural, that can figure in an empirical science of humanity. My argument for this thesis is transcendental: As planners, capable of agreement or disagreement in plan, we are each committed to this naturalistic-sounding thesis [...]: There is a natural property that constitutes being what one ought to do. [...]

And what property is this? The question is not linguistic; rather, it is the grand, basic question in ethics, the question of how to live. You accept an answer to this question if you have fully thought out what to live for and have come to a conclusion. (Gibbard 2002, 54)

Gibbard may be right that plans for action must be “couched in empirical, naturalistic terms” (ibid., 57) in order to be useful and to be objects of agreement or disagreement. However, this does not mean that *being the right plan* is a natural property. More generally, what one ought to do in a given situation can presumably be specified in purely natural terms, but that doesn’t mean that the property of being what one ought to do itself must also be a purely natural property, on a metaphysical par with the redness of a house or the height of a tree. The latter idea simply seems like a category mistake.

In defense of this rather harsh claim, consider what it would be like, on naturalism, to become convinced that the theory I defend in this thesis is correct. Assume that we have resolved the questions discussed in chapter 3 and agree on the natural properties of pleasure and pain. Thus, you have a clear grasp of the natural property of being pleasant, and you know some experiences which have this property. When you come to believe that hedonistic egoism is true, you become convinced that this property just *is* the property of being what there is a reason to seek. You do not change your mind about the property itself; there is, after all just one property here, i.e. the natural one. What happens is, rather, that you realize that your concept of an authoritative reason refers only to this natural property.

This is not the reaction I intend my thesis to provoke. My aim is not to point to a property of certain experiences that we can all agree is there and then claim that this property is what our normative discourse is ultimately all about. The view I defend is supposed to be about properties, not concepts: I want to say something about the world, and not about how we talk. More precisely, I want to say that the experiences in question have a property that most philosophers do not think they have and that is very different from the uncontroversial natural properties, namely that of being the only thing that matters for its own sake.

My present objection is not that the goodness and phenomenal quality of pleasure could not possibly turn out to be one and the same property. In section 8.3, I will suggest that they very well may be. What I want to claim is that, if this is indeed the case, then the single property in question would not be a straightforwardly naturalistic one. It would have to be a

property that is very, very special: a feeling that, by its very nature, demands to be sought. While I do not think positing it *conflicts* with anything science tells us about the natural world, it would hardly be just more of the same, either.

I believe, then, that if hedonistic egoism is to be correct as a theory of ultimate reasons, it would either have to be the case that the way pleasure and pain feel defy naturalistic description, or that, if such a description can be given, these states also have the further, non-natural properties of being intrinsically good and bad.⁹⁸

These remarks about naturalism have been very brief, and there are of course more sophisticated versions of the doctrine to be found in the literature. Since a thorough treatment of normative naturalism is beyond the scope of this thesis, I will just refer the reader to Parfit's discussion of naturalism in his forthcoming book *On What Matters*, where similar and further objections to the doctrine are discussed in great detail (Parfit forthcoming, ch. 24-6).

There is, as far as I know, no non-naturalist way of demonstrating that the queerness of independent normative authority is only apparent. One could take the edge off it a bit though, if it could be shown that a similar queerness is found in other areas of inquiry as well. This strategy is often suggested and sometimes pursued in contemporary philosophy. For instance, Russ Shafer-Landau argues that the "intrinsic normativity" of ultimate reasons is no more mysterious than the intrinsic normativity found in epistemology (Shafer-Landau 2003, 205-9). Here, too, there are reasons that seem to be non-derivative and desire-independent. As Shafer-Landau observes, "there is a reason to think that two and two are four—the fact itself provides one with reason to believe it. One needn't show that such belief is somehow related to one's adopted goals in order to justify believing such a thing." (ibid., 206) Since not even error theorists are prepared to give up intrinsic normativity in epistemology (how could they then claim to have the "best" theory?), Shafer-Landau thinks there is no reason to be sceptical of intrinsic normativity in ethics, either. Jean Hampton makes a similar argument based on the objectivity of scientific standards (Hampton 1998, ch. 6).

The problem with this response to the argument from queerness is that, as explained in section 2.1, reasons for actions are different from reasons for belief. To say that there is an "intrinsic" epistemic reason to believe that two and two are four is to say (roughly) that

⁹⁸ David Brax claims that his account of hedonism as "the explanation of value" is perfectly naturalistically respectable, and so not vulnerable to the argument from queerness. Brax may be right about this, but his form of hedonism is not intended as a theory of ultimate reasons. In fact, his claim that pleasure is valuable seems to mean little more than that we are motivated to seek it (Brax 2009, see in particular 201-2).

merely understanding the proposition provides one with evidence that it is likely or certain to be true, in a desire-independent way. It is *not* to say that one just has to *do* what (if anything) it takes to attain or retain this belief. Facts about evidence may be objective facts, but that is not the same as saying that there is an ultimate reason to be concerned with them. Shafer-Landau never offers any positive argument for thinking that there is; he just lumps the two kinds of reasons together under the heading of “intrinsic normativity” and assumes that they must be equally (un)acceptable to a naturalistically minded philosopher. I suspect that most such philosophers would find ultimate reasons much queerer than objectivity in epistemology. However, the important point in this context is that these two issues are distinct and must be settled on their own merits.

Epistemological standards *could* be formulated in terms of reasons for action. They would then inform us (again, roughly) that if we want to understand the world, we have to do such and such (see Tännsjö 2010, ch. 8 for a much more detailed account of the epistemic goal). This way of formulating epistemological claims seems unnecessary and a bit odd, since forming a belief is, at least typically, not a voluntary action at all. More importantly, there are of course no ultimate reasons to be found in this form of epistemology, since the reasons for action in question lack independent authority: their normative force would be dependent on an ultimate reason to seek understanding. Claiming that there is a reason of the latter kind would simply beg the question against the anti-realist.

Finally, one could interpret epistemology as a direct attempt to provide ultimate reasons for action. If so, it wouldn't just tell us what to believe if we want true or justified beliefs; it would tell us that we just have to believe certain things, period. Terence Cuneo starts from this conception of epistemology in his attempt to show that moral realism is no more vulnerable to the argument from queerness than epistemological realism (Cuneo 2007, 58-61). Given this starting-point, it is not particularly difficult for Cuneo to reach his conclusion (see *ibid.*, 92-8). The question is, of course, why we should accept the starting-point. Cuneo appears to do so because he thinks that alternative conceptions will lead to implausible forms of epistemological relativism or nihilism (*ibid.*, ch. 4-7). However, it is by no means clear why, if we think that epistemology aims to provide instrumental reasons, these reasons must differ from person to person (*ibid.*, 188-9). More importantly, Cuneo never considers the possibility that epistemology isn't about reasons for action at all, but about reasons for belief. As explained in section 2.1, reasons for belief may be perfectly objective even if they do not entail any reasons for action. In order to hold that, say, the theory of evolution is better epistemologically justified than creationism, we do not have to claim that there is an ultimate

reason to have correct beliefs about the origins of the species. We could just think that it is, given the evidence, more likely to be true.

Surely it is this notion of epistemology without ultimate reasons that philosophers who find the argument from queerness convincing will want to embrace. Those who think they can avoid the argument from queerness by relying on an analogy with epistemology need to explain precisely what is wrong with this understanding of epistemology. I am not aware of any author who has done so.

Until now I have been discussing the ontological queerness of ultimate reasons: the problem has been that they seem to be very strange things. As Mackie stresses, however, there is also a corresponding epistemological queerness: by what mechanism or capacity do we get to know objective values? As we have seen, it is not enough to be acquainted with the natural properties of the things that have value; we would also have to have access to the further non-natural fact that they are valuable. Assume, for instance, that the philosophers mentioned in section 6.3 are right that friendships are sources of ultimate reasons. I believe that I have a pretty good idea of what the descriptive properties of friendship are like. I know that one enjoys being with one's friends, that one automatically thinks of how one can help them out when they are in trouble, that it is hard to break off a friendship, that it is psychologically healthy to have someone to confide in, and so on. However, I'm just not able to see that friendships are, in virtue of these things, an independent source of ultimate reasons. No matter how hard I try and no matter how much I like my friends – I find that the independent authority is not there. If it is true that friendships have or generate this kind of authority, I must in some way be epistemologically deficient. But how, exactly?

Few contemporary philosophers would claim that there is a clearly separate intuitive faculty that delivers simple and infallible pronouncements on normative questions. Rather, the idea is that we have intuitive reactions to claims about reasons which are of many kinds and on various levels of generality, with only prima facie epistemological authority (see e.g. Daniels 1979; Audi 1997; Rawls 1999b, ch. 15, sect. II). In the coherentist reasoning that follows, we can rely on much the same argumentative principles as we do in other kinds of inquiry. However, we must not forget that there must at some point be a distinctively normative input to this process. Normative observation may not be analogous to visual perception, but it must be some sort of epistemological capacity or other. Knowledge of ultimate reasons cannot simply appear out of nowhere. Appealing to coherentist reasoning does not solve the problem, but only postpones it. In order to see this, imagine a group of

psychics who claim to have direct knowledge of events that occurred hundreds of years ago. This epistemological capacity would, of course, be extremely queer. It is not as if it becomes any less mysterious if the psychics explain to us that they do not have a separate faculty that perceives past events, that they have convictions about the past on various levels of generality, or that they are willing to revise any particular judgment in order to generate a maximally coherent story.

We need, then, an informative account of how we get in epistemological touch with ultimate reasons. It would be unreasonable to hold that until such an account is provided, realism about reasons is completely untenable. As an analogy, the fact that it would be somewhat unclear how we get to know realist mathematical facts is hardly a decisive objection to mathematical realism.⁹⁹ However, the epistemological obscurity certainly doesn't help realism, either. It gives us at least some reason for being skeptical of the notion of ultimate reasons, but what our final judgment should be depends on many other considerations.¹⁰⁰

7.2 Disagreement

People living in the middle ages could reasonably be realists about the physical world even if they were not able to explain what kind of entity it is and how we get in touch with it via perception. Note, however, that this is only the case because humans tend to have very similar perceptions. If there had been a lot of disagreement on basic perceptual questions, there would have been much less basis for confidence in the existence of an objective physical world. In normative philosophy, this kind of disagreement is precisely what we find. There is – literally, I think – no substantial claim about normative reasons for action that would not be disputed by at least some philosophers. According to Mackie, this gives us additional reason to be skeptical of the existence of objective values (Mackie 1977, 1.8).

⁹⁹ Parfit puts great weight on this analogy when discussing the epistemological queerness argument (Parfit forthcoming, ch. 29-30; see also Crisp 2006, 3.3).

¹⁰⁰ It should be mentioned that there are philosophers who attempt to derive substantive claims about ultimate reasons without relying on substantive normative intuitions. The arguments by Kant, Korsgaard and Gewirth discussed in earlier chapters exemplify this approach. If an argument of this kind could be made to work, objective value could be established without any epistemological queerness (and perhaps also without any ontological queerness). As explained before, I think these arguments won't work, and I am proceeding on the assumption that they don't.

Not all disagreement about reasons and values poses a threat to normative realism. First, some of the disputes that seem to be about normative issues are in fact mainly about factual issues (see e.g. Brink 1989, 202-3). Second, realists need not worry about the strongly diverging views among all those people who are not particularly intelligent or open-minded, or who haven't spent much time pondering questions about ultimate reasons. Third, even talented and well-educated philosophers sometimes go wrong in their reasoning: they talk past each other, make too hasty judgments, conduct thought-experiments in the wrong way, make mistaken inferences, and so on. To the extent that the disagreement in normative philosophy can be attributed to these kinds of error, it clearly gives us no reason to be skeptical of realism.

However, I do not think that all the disagreement among moral philosophers can be explained away in this manner. A significant part of it stems from another source, namely differences in normative intuitions. One may, as we saw in section 5.2, think that normative facts can be discovered without appeal to substantive normative intuitions, and then no amount of intuitive disagreement could undermine one's commitment to realism. But if, like most philosophers concerned with ultimate reasons, one thinks that intuitions play a crucial role in normative discovery, the prevalence of intuitive disagreement poses a problem. It is this form of disagreement I will be discussing in the remainder of this section.

When realists respond to the objection from disagreement, they tend to interpret it in the following way:

(1) If realism were true, there could be no plausible explanation of why there is so much disagreement.

They then go on to argue that there are various explanations that are compatible with realism and that ethics is, as Nagel puts it, "an area in which one would expect extreme variation of belief and radical disagreement however objective the subject actually was." (Nagel 1986, 148; see also Brink 1989, 7.4; Shafer-Landau 2003, ch. 9) That may be correct. However, I think there is a more fundamental challenge to realism to be found in the argument from disagreement. The objection should be read like this:

(2) Perhaps we can account for all disagreement in a way that is compatible with realism. But if there is so much disagreement, why believe in realism in the first place?

Presumably, the main motivation behind realism about ultimate reasons is an impression that some particular convictions about ultimate reasons really are correct. It is difficult to see how a person who is not willing to assent to any substantive value claim could at the same time find value realism plausible. But if our normative beliefs are based on intuitions and if intuitions are as unreliable as they have to be for intuitive disagreement to be compatible with realism, then what underlies the initial confidence in these beliefs? Why think there is anything behind the appearances?

In the following, I will consider some possible replies. First, it could be claimed that explanation for why philosophers give contradicting answers to normative question is not first and foremost that they have different intuitions, but rather that philosophical theorizing is underdeveloped. The idea would be that prevalent intuitions do, in fact, support one and the same outlook; it is just that this outlook hasn't yet been developed. Once (or if) it is, we can all, at least after some coherentist adjustments, acknowledge it as intuitively acceptable. Currently, there is, of course, no such outlook in the literature. In the years that have passed since Derek Parfit wrote that non-religious ethics is in "such an early stage" that "it is not irrational to have high hopes" (Parfit 1984, 454; see also Brink 1989, 206-8), it has become more and more difficult to imagine what it could be like. (Parfit himself makes a heroic attempt to provide one in his forthcoming book *On What Matters*.) Ethics has made progress in the sense that we have a larger pool of theories to choose from and a better understanding of what each theory implies, but not, as far as I can see, in the sense that there is more agreement on substantive issues. While it is certainly possible that there exists an outlook that is uniquely favored by the intuitions of contemporary philosophers, it does not seem very likely. It seems much more likely that we simply have different intuitions.

Second, realists often stress how normative judgments are subject to particularly strong biases, since they concern issues that are important to us and provoke distorting emotional reactions (Nagel 1986, 148; Shafer-Landau 2003, 219). Perhaps the solution to the problem we have been considering is to focus on intuitive judgments that are made "in a cool hour", after considering the matter from all relevant perspectives, and by people who have no unusual interests or experiences. In fact, it seems to be a working assumption in much of the literature that if we only remove all specific factors producing a bias, the result will be intuitive judgments that converge and can be trusted.

An explicit formulation of this assumption can be found in John Rawls' first paper, "Outline of a Decision Procedure for Ethics" (Rawls 1999b, ch. 1). His idea is that our ethical principles can be based on pre-theoretical normative judgments, as long as we make sure that

(among other things) the judgments are made by persons who have thought carefully about the matter, are not personally involved, will not be punished or rewarded for their judgment, are open-minded, familiar with general social facts, capable of sympathy, fairly intelligent, and willing to use inductive logic, to search for reasons, and to take into account his own biases and prejudices. (ibid., 1-7; see also Rawls 1999a, 42). Thus, if people are just reasonable and conscientious, they will eventually get their intuitive judgments right.

The problem is that even reasonable and conscientious philosophers seem to have different intuitions. To cite a well-known example, G. E. Moore thought he could see, when he thought about it carefully, that there is an ultimate reason to bring about beauty (G. E. Moore 1903, 126-8, 144-7, 236-48; see also Regan 2003, sect. V). On the other hand, he did not, as far as I know, accord the same normative status to promise-keeping. W. D. Ross thought, on the same intuitive grounds, that while there is no ultimate reason to promote beauty (Ross 1930, 130-1; see also Brink 1989, 219), there is an ultimate reason to keep promises (op. cit., ch. II). Did Moore's and Ross' intuitive judgments violate any of Rawls' requirements? I doubt it. More generally, it is implausible to hold that when two professional moral philosophers have different intuitions, this is typically because at least one of them is not thinking clearly or is subject to some special bias of the kind mentioned above, and that when these errors are corrected, agreement ensues. This response to the argument from relativity underestimates how good moral philosophers are at what they are doing.

One could hold that the biases making intuitive normative judgments go wrong are so subtle that one needn't notice them even if one is a reasonable and conscientious professional philosopher. However, as long as this claim remains on a vague and general level, it should undermine one's confidence in all intuitive judgments – including those that constituted one's original basis for being a moral realist. If this kind of reply is going to work, one would have to give an account of the mechanism behind the bias, explain how it fails to affect at last some intuitive normative judgments and show that these intuitions are widely shared. This is, in fact, precisely what I will attempt to do in the next chapter. For now the important point is that one must have some story of this kind to offer in order to square a belief in realism with the existence of persistent intuitive disagreement. It is not enough to observe that intuitive normative discovery could be generally difficult.

Third, some intuitive judgments, typically concerning particular cases, are of course almost universally shared. It is tempting to think that the argument from disagreement fails because it overlooks this fact. For instance, Torbjörn Tännsjö writes that

while there is much disagreement about speculative moral principles, intended to ‘explain’ particular moral judgments, [...] there is also much agreement. There are indeed moral facts, and, if we succeed in taking up an impartial stance, we are sometimes capable of recognizing them. All decent and impartial people, who consider the question, would agree that it is wrong to torture an innocent child, at least unless there are any pressing reasons of any sort to do so. (Tännsjö 2010, 50-1; see also Shafer-Landau 2008, 92-4)

I actually don’t agree that torturing an innocent child is wrong in Tännsjö’s sense, since he takes claims about wrongness to be claims about what I in this thesis call ultimate reasons. I do, however, have a clearly negative intuitive reaction towards the idea of torturing children. If virtually all philosophers have this reaction, wouldn’t a realist judgment based on it be immune to the argument from relativity?

Assume first that we are able to make a principled distinction between those intuitions that are widely shared and therefore trustworthy, and those that are not. Perhaps they have a unique causal history, stem from a special cognitive or neural mechanism, or involve a distinct phenomenological feel. Then we would have good reason to deny that the argument from disagreement succeeds as an objection to all forms of normative realism. I will, as I have said, suggest and defend such a distinction in the next chapter. However, defenders of normative realism do not typically attempt to do this. They are content to just point out that certain intuitive reactions are very common. Perhaps they are also typically stronger – though it should be kept in mind that Moore’s and Ross’ intuitions also seem to have been quite robust. If this is all that sets them apart from more controversial intuitions that cannot be appealed to in a defence of normative realism, then the worry remains that they are ultimately just more of the same. The suspicion towards intuitions about beauty or promise-keeping seems to generalize, and one is left wondering whether the intuition that there is a reason not to torture children really must reflect an objective normative fact after all. The intuitions seem different only in degree, while what would save the realist position (at least from the present objection) is a difference in kind.

As an analogy, imagine that you are struck by the extensive and persistent disagreement on aesthetic questions and takes this to pose a threat to aesthetic realism. An aesthetic realist then points out that virtually everyone thinks that sunsets are very beautiful. As long as this aesthetic judgment does not differ from other aesthetic judgments in any other way than by being made with more force and by more people, it seems to be of quite limited help in dealing with an aesthetic version of the argument from relativity.

A realist could point to cases like Tännsjö's not in order to show that he need not worry about the disagreement about other cases, but in order to show that the anti-realist has a problem. If, from an anti-realist point of view, it would have to be just a freak coincidence that nearly everyone has the intuition that one ought to refrain from torturing children, then anti-realism wouldn't be a tenable alternative. I will explain how an anti-realist can account for normative agreement in the next section (see also section 8.2).

It seems, then, that persistent intuitive disagreement among reasonable and conscientious professional philosophers does put realism about reasons in a somewhat questionable light. According to Shafer-Landau, we will, if we accept this line of reasoning, be lead to anti-realism in all areas of philosophy. There is, after all, deep and intractable disagreement in all philosophical disciplines. This general form of anti-realism is a position few philosophers would be willing to accept. The lesson Shafer-Landau draws is that we should reject the argument from disagreement: "If we are reluctant to see all metaphysical or epistemological opinions as nothing more than expressions of personal commitments, then we ought to be equally reluctant to take persistent ethical disagreement as a good basis for inferring the truth of moral antirealism. The disagreements that arise within moral discussions are not probative evidence for an antirealist diagnosis of morality." (Shafer-Landau 2003, 220)

If the argument from disagreement had been intended to be decisive on its own, the analogy with other disciplines would have been an excellent reply. However, this anti-realist argument, too, should only be understood as a piece of a larger puzzle. It puts some pressure on realism, but whether we ought to reject the outlook depends on what the realist has to say about other matters. Importantly, it also depends on what the alternative would be like. Is anti-realism a credible position?

Once the argument is qualified in this way, I think it is clear that it does apply to other areas of philosophy. The fact that philosophers do not manage to agree on whether we have free will or whether physicalism is true *should* make us wonder whether there really is any truth there to be found. In most of these other areas, however, anti-realism has little initial plausibility. Most philosophers find it hard to understand, for instance, how all claims in philosophy of mind could be meaningless or false. The same goes for metaethics: surely there must be a truth about the philosophical status of ethics that is independent of what we happen to take or want it to be.

When it comes to first-order normative claims, however, many philosophers have thought that anti-realism is a perfectly understandable and even fairly plausible view. If it is, then we should, given the difficulties surveyed so far, be prepared to abandon realism. In the next

section, I give an outline of what I consider to be the most promising version of anti-realism about reasons.

First, however, I would like to point to a consequence of the disagreement I have described in this section that is quite straightforward, but appears to be surprisingly often overlooked or disregarded. While an absence of a systematic explanation of how normative intuitions go wrong does not completely rule out the possibility that there are objective normative facts, it does prevent any controversial intuitively based first-order view from being *justified*. The problem is not the mere fact that other people disagree with what we are saying, but that we cannot give any plausible explanation of why they do so that doesn't at the same time undermine our confidence in our own views.

As an example, how could Ross explain what he took to be mistakes in Moore's intuitive judgments? Moore could not reasonably be accused of being a sloppy or unreasonable thinker, so the problem would have to be that it is just generally very difficult to get one's intuitive judgments right. But if this is true, what then of Ross' own intuitive judgments? It seems that, unless Ross is able to demonstrate somehow that he is a better intuitive knower than Moore, he is not justified in trusting his own intuitions. Though it may still seem to him that he is right, and though he might continue developing his theory, he should suspend judgment.

To illustrate the problem further, consider William K. Frankena's list of what philosophers have taken to be intrinsic values:

Life, consciousness, and activity
Health and strength
Pleasures and satisfactions of all or certain kinds
Happiness, beatitude, contentment, etc.
Truth
Knowledge and true opinion of various kinds, understanding, wisdom
Beauty, harmony, proportion in objects contemplated
Aesthetic experience
Morally good dispositions or virtues
Mutual affection, love, friendship, cooperation
Just distribution of goods and evils
Harmony and proportion in one's own life
Power and experiences of achievement
Self-expression

Freedom

Peace, security

Adventure and novelty

Good reputation, honor, esteem, etc. (Frankena 1973, ch. 5)

This is a very inclusive and open-ended list. It is difficult to think of a candidate for intrinsic value that could not be made to fit into one of the categories above. Whatever your convictions are about which of these things that belong on the correct objective list of intrinsic values, most philosophers will have differing intuitions. Why should you think that you of all people happen to be right? Are you simply a better intuitive knower than other moral philosophers?

My present point is not that we should stop according a general weight to intuitions when doing normative philosophy; there may not be any better alternative. Rather, it is that authors who make intuitive claims with realist pretences about reasons or values are too confident. They should at least admit that they are fumbling around in the dark. Thus, I agree with Roger Crisp's remarks in *Reasons and the Good*:

usually, when confronted by someone with a different ethical viewpoint from ourselves, we are inclined to think that they have failed to see something that we have appreciated. But here we should remember that they will think the same about us, and the question is whether either of us is justified in thinking that the other is in a worse epistemic position. I suggest that such justification in ethics is quite rare, and that to this extent philosophical ethics is characterized by an unjustified dogmatism or over-confidence. (Crisp 2006, 94)

When you are being asked to believe in a normative theory, you are being asked not only to accept a view of the world where the theory is true, but a view of the world where the theory is true and most people are mistaken. We should not forget how important the latter part of the story is. A theory of ultimate reasons that is able to explain why the intuitions that conflict with the theory can be disregarded in a way that doesn't undermine the credibility of the intuitions that support it would possess a very rare virtue indeed.

7.3 How there is an apparently plausible alternative

In this section, I will sketch how we could attempt to make sense of the world without appealing to ultimate reasons. It is obvious that we do not need ultimate reasons in order to account for phenomena such as earthquakes, cancer or perception. The decisive question is whether the anti-realist can explain why there are so many people who believe in ultimate

reasons and act on what they take to be ultimate reasons without at any point positing that these reasons actually exist.¹⁰¹ Unless there is a credible account of how the realists go wrong, anti-realism, too, demands that we accept mysteries.

On the other hand, if we do have such an account, then realism seems to have little going for it. There is no justification for populating one's ontology with dubious entities that serve no explanatory function. Just as we can discard the belief in curses because we no longer need them to explain any misfortunes, we can discard the belief in objective value if we come to see that there are biological, psychological and sociological explanations of why human beings make reasons judgments and engage in practices based on them. Similarly, there is no rational justification for believing in a god if we have a plausible explanation for everything that goes on in the world, including sincere avowals from religious people that the god really does exist, without at any point appealing to the god in question.

It is important to note that the present question is not whether normative facts have independent causal power. If that had been the relevant test, mathematical facts and epistemological standards would presumably have had to be banished from our ontology as well. The crucial question is whether a given entity is necessary for making sense of what happens in the world. Mathematical facts are necessary in this sense, since many of our best scientific theories are formulated in mathematical terms (see Harman 1977, 8-9). So are epistemological standards: without them, we could not explain anything at all.¹⁰²

Also, there is no need to insist that ultimate reasons *cannot* exist unless they are presently part of the best explanation for what we experience. Surely it is conceivable that they do; explanation is an epistemological notion, not an ontological. The point is just that we don't have any reason to *believe* that they exist if they are not needed for explaining anything.

Gilbert Harman famously made a similar claim about moral facts (Harman 1977, ch. 1-2). However, realists tend (mistakenly, perhaps) to interpret him as saying that the explanatory irrelevance of moral facts can be demonstrated a priori: moral facts are just not the kind of

¹⁰¹ I do not want to speculate on exactly how common the belief in ultimate reasons is. It may be that value realists are a minority, but they are at least a significant minority, comprised by (among others) some very intelligent and reflected people.

¹⁰² When Shafer-Landau defends moral realism against the present objection, he quickly slides from discussing an "explanatory requirement" to a "causal requirement" (Shafer-Landau 2003, 98-100), which he rightly finds implausible when generalized, since it implies that we would also have to reject our belief in epistemological standards and so have no basis for saying that anti-realism is more plausible than realism in the first place (ibid., 113-4).

things that could ever be a part of the best explanation for any phenomenon (see Sturgeon 1988; Brink 1989, 7.3). As the realists go on to show, this claim seems to simply beg the question. We are never told why we have to accept that moral facts cannot play any explanatory role.

This is not the anti-realist argument I am presenting here. I doubt that belief in ultimate reasons can be dismissed as a violation of some epistemological standard right from the outset. It seems an open question whether ultimate reasons are part of the best explanation of what happens in the world: in order to answer it, we need to develop competing realist and anti-realists accounts of value practices and see which outlook that, in the end, appears more plausible.

In the remainder of this section, I will outline a projectivist explanation of the prevalence of the belief in realism. This is not the right occasion for going into details, and many of the details remain to be filled in by further research in any case. My aim is only to show that it does not seem unlikely that some version of the projectivist outlook could be correct. It is, to some extent at least, empirically supported and continuous with the sciences, and it is not vulnerable to the arguments from queerness and disagreement.

I will begin by mentioning a few evolutionary considerations. I should stress that I do not intend these considerations by themselves to constitute an objection to realism (see Parfit forthcoming, ch. 29-30). It makes perfectly good evolutionary sense to claim that the belief in ultimate reasons is an epistemologically respectable by-product of our general capacity for reasoning, which has certainly been conducive to reproduction. My point is more modest: the idea that human beings have ended up believing in ultimate reasons even if there aren't really any such things *also* makes evolutionary sense.

I am not aware of any good reason to think that specific normative beliefs, such as the conviction that equality is intrinsically valuable, could have been favored by natural selection (see Prinz 2008). However, it may be that a general propensity to form beliefs about ultimate reasons has been. Human likes and inclinations are notoriously unstable when it comes to both their content and their motivational power. As Richard Joyce points out, there seems to be an evolutionary advantage in having the capacity for a more permanent motivational structure, i.e. in being able to form *commitments* (Joyce 2006, 4.2-4.3; he is talking about moral commitments, but the point presumably generalizes). Commitments presuppose, or are at least facilitated by, beliefs that their object is somehow good or worth having. In particular, this is true if the goodness is perceived as being in the object itself, rather than merely a

creation of our minds. As an example, consider the common reluctance to cheat during cooperation, even in cases where there is no risk of getting caught. If we merely *liked* to avoid cheating, in the same way that we like sex or food, then the motivation to play by the rules would easily be overcome, and the evolutionary benefits provided by reciprocity would remain unavailable. But if there is a belief that cheating is somehow inherently bad, or that the rules have some sort of “to-be-followed-ness” built into them, then a robust commitment to play by the rules becomes possible.

One cannot always provide an evolutionary explanation of why we hold a class of beliefs without assuming that the beliefs are true. For instance, having perceptual beliefs is clearly conducive to reproduction, but this is so only because they tend to be correct. Similar considerations apply to our ability to form generalizations on the basis of experience and to do mathematics and logic. In the case of beliefs about ultimate reasons, however, it is hard to see how they would in any way have to be reflections of objective truths in order to fulfill an evolutionary function (unless of course one takes questions about ultimate reasons to be reducible to questions about what is conducive to reproduction, but this is not exactly the standard view). It thus seems that we can make sense of the prevalence of the belief in ultimate reasons from an evolutionary point of view without having to postulate that the objects of the beliefs are real.

The fact that we can tell some evolutionary story about a purported phenomenon does not mean that it actually has evolved. If the idea that the belief in ultimate reasons is some kind of massive collective illusion is to be at all plausible, then we must have both (1) an account of the mechanism that generates the illusion and (2) some empirical data showing that human beings are in fact equipped with this mechanism. As regards the mechanism, the only plausible alternative is, as far as I can see, that normative judgments are a result of *the projection of subjective responses*. When we respond positively or negatively to some object, the positivity or negativity comes to be represented as belonging to the object itself. For instance, our admiration for knowledge is experienced as some sort of perception that knowledge is *admirable*, and the aversion to hurt innocent people is experienced as some sort of perception that doing so is inherently forbidden.

It is important to stress that projectivism is a theory about the causes of judgments about reasons, and not about their meaning or content. It must not be confused with the emotivist thesis that claims about reasons are merely expressions of emotions. Projectivism accepts that we (or at least some of us) *purport* to describe genuine normative properties when we talk

about reasons. It is just that there aren't really any such properties in the world. We are fooled into believing in them by the phenomenological nature of our subjective responses.

Also, a projectivist will not deny that there is such a thing as reasoning about value, or that we often can provide rational justifications of our normative beliefs. What he claims is that the beliefs are not (at least not in the standard case) caused by the reasoning. Rather, the reasoning is ultimately just rationalization: post hoc justification of beliefs held on intuitive or emotional grounds, provided for the sake of appearing rational to oneself or others. For instance, those who have thought that guilt arises as a result of a realization that one has done something wrong have put the cart before the horse: it is because we feel guilt that we end up finding some reason for saying that what we did was wrong.

This outlook has, of course, a long tradition within moral philosophy (see e.g. Hume 1743, 88-90; Nietzsche 1887, III.16; Mackie 1977, 1.10; Blackburn 1993, part II). It is also the outlook Joyce relies on in his evolutionary debunking of morality: since emotions have been around for a long time, they can be utilized in an account of how the propensity to believe in ultimate moral reasons was favored by natural selection (Joyce 2006, 4.4). Until recently, projectivism was mainly based on philosophical speculation. In the past few decades, it has, however, received apparent support from a growing body of psychological research.

First, there is a general finding that human beings are often influenced by factors that are normatively irrelevant when they evaluate, act or choose in everyday life. To name a very simple example, when asked to choose between various products, people tend to prefer those that are placed to the right (Nisbett & Wilson 1977). Or consider a study cited by Joshua Greene where the participants were presented with the opportunity to give money to strangers (Greene 2008, 49-50). Some were told that the identity of the stranger had already been determined by a random mechanism, while others were told that it would be after they had made their choice. Obviously, this difference cannot be normatively relevant. However, it turned out that those who had already been assigned a recipient gave on average 60% more.

Many more examples of this phenomenon could be given (see e.g. Gigerenzer 2008; Sinnott-Armstrong 2008). What is particularly interesting about the findings is that human beings often have little or no introspective access to the mechanisms responsible for generating their judgments and actions. It *seems* to us as if we are guided by what we take to be the best reasons, and if asked why we do what we do, we almost always come up with some rational-sounding explanation. No participants said things like "I just prefer products that are put on the right side of other products" or "I subscribe to a principle according to

which it makes a great difference whether one can help a stranger who was selected randomly a few minutes ago or a stranger who will be randomly selected in a few minutes.”

Psychologists believe that the explanation for this phenomenon is that human beings are equipped with two fairly distinct mechanisms for processing information. First, there is the capacity for explicit reasoning that is unique to (or at least uniquely developed in) humans. This capacity is epistemologically reliable, consciously accessible and under rational control in the sense that we can decide what to think about and when to do it. The downside is that it is slow and very demanding in terms of attention. This is why we, in everyday life, mostly rely on a phylogenetically older system for producing judgments or evaluations. This system is fast, automatic and very undemanding in terms of attention. Also, we have no direct rational control over the content of the responses, or even whether to give an intuitive response at all. Accordingly, automatic responses tend to persist in the face of explicit cognitive convictions that they must be mistaken. A sad example of this phenomenon is the fact that people who are convinced that there is no rational basis for racism still tend to display unconscious negative reactions towards people of other races. Finally, automatic judgments, especially those concerning values, are typically emotionally or affectively laden, in a way that conclusions established by explicit reasoning usually are not.

Although we do, of course, experience the judgment or evaluation that the automatic system generates, we have no conscious access to how it is reached. What we have conscious access to is only our explicit reasoning concerning our conduct. Thus we tend to think, falsely, that the reasoning is the cause of our actions, when in fact it is often nothing more than rationalization.

The automatic system performs many tasks astonishingly well. Consider, for instance, the hunches we sometimes get that something is wrong with a friend. Even if we have no conscious access to any facts or observations that could justify them, they quite often turn out to be correct. However, since it is not under any kind of direct rational control, the automatic system is also susceptible to various biases and distortions illustrated by the examples above. It has evolved in order to give fast and useful responses in the situations most commonly faced by the organism, and not to accurately track the truth. As Paul Slovic puts it, “appears at once both wondrous and frightening: wondrous in its speed, and subtlety, and sophistication, and its ability to “lubricate reason”; frightening in its dependency upon context and experience, allowing us to be led astray or manipulated—inadvertently or intentionally—silently and invisibly.” (Slovic 2002, 1349)

Even if we are to a great extent guided by the automatic system in everyday life, it could still be the case that we rely on conscious reasoning when asked to explicitly reflect on normative questions. The research into this question has, as far as I know, focused almost exclusively on judgments about moral reasons. However, it seems reasonable to suppose that the results can be generalized to other kinds of reasons for action.

A few decades ago, moral psychologists tended to give a very rationalistic picture of how adults form moral judgments. More recently, this picture has been challenged, most prominently by the findings of Jonathan Haidt and his colleagues. In one of the experiments, the participants were asked to read the following story:

Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that, was it OK for them to make love? (Haidt 2001, 814)

Most of the participants immediately judged that it was not. When asked for reasons, they typically pointed to the risk of inbreed or the damage done to the relationship between Mark and Julie. But of course it is stipulated in the story that there is no such risk or damage, so that cannot be the reason why what they did is wrong. Eventually, the participants ran out of rational explanations for their judgment. If the judgment had been made on the basis of principles or reasons, one would expect it to be changed or withdrawn. But this is not what happened; most of the participants ended up saying things like “I don’t know, I can’t explain it, I just know it’s wrong.” (ibid.) On the basis of this and many other experiments, Haidt concludes that automatic emotional response, rather than arguments and principles, is the driving force behind moral convictions. In other words, “the reasoning process is more like a lawyer defending a client than a judge or scientist seeking truth” (ibid., 820; see also Greene 2008).

If this picture is correct, it is easy to see why deep moral disagreement would be both common and persistent. People have different intuitive reactions, and attacking the arguments with which other people justify their moral views is bound to have little effect, since the views are not held on the basis of arguments in any case. As Haidt puts it, moral arguments would be “like shadow-boxing matches: each contestant lands heavy blows to the opponent’s shadow, then wonders why he doesn’t fall down.” (op. cit, 823)

Of course we should not forget that there is also much moral agreement. In fact, it could be plausibly claimed that there has been a significant convergence on many normative questions the past few centuries, at least in the Western part of the world. For instance, we now agree that racism and sexism is unjustifiable, and that governments should protect the basic rights and liberties of every citizen. This development might seem to contradict the claim that people are not susceptible to moral argument, and some realists have taken it as support for their view (Nagel 1986, 148; Brink 1989, 208-9). However, arguments are not the only possible source of convergence. According to Haidt's "social intuitionist" model of moral judgment, the automatic system is, over time, heavily shaped by socialization (Haidt 2001). We emulate the moral views of those around us in much the same way as we emulate their language or sense of style. A general cultural convergence will, therefore, also lead to moral convergence even if there are no objective moral values. The fact that almost everyone opposes slavery these days no more means that it is objectively wrong than the fact that more and more people are wearing jeans and speaking English means that we are approaching an objectively ideal garment or language.

Until now I have been discussing how ordinary people form normative beliefs. But what about professional philosophers? Do they just offer more elaborate rationalizations for their automatic intuitive responses, or are they able to make claims about ultimate value relying only on some more epistemologically respectable capacity?

There is no doubt that philosophers are now and then willing to override their intuitive responses when they conflict with what they take to be plausible general principles. But this does not mean that they are beyond projectivist criticism, since the general principle may, in turn, be intuitively supported. As an example, consider Robert Nozick's book *Anarchy, State and Utopia*, which is full of ingenious right-based arguments for conclusions that are in no way intuitively obvious. But how does Nozick justify the claim that human beings have rights in the first place? He simply starts his book by stating that "individuals have rights and there are things no person or group may do to them." (Nozick 1974, ix) And this judgment seems easily susceptible to a projectivist analysis. Nozick – like most of us – has grown up to develop a tendency to make automatic and emotionally laden judgments concerning what one may or may not do to other people and to feel resentment or indignation when these intuitive rules are broken. These emotional responses are then projected onto the world: it comes to seem to Nozick as if human beings possess something called individual rights that by their very nature demand to be respected. The projection is reinforced by living in a culture where

rights have paramount political importance and where other people engage in realist discourse about rights all the time.

Though I am not, obviously, in a position to fill in the details in the projectivist explanation of Nozick's moral views, it does not seem too far-fetched to suppose that there is such a detailed explanation to be given. Also, it would seem that we could find similar appeals to automatic emotive cognition at crucial stages in any other argument for a substantive normative position.¹⁰³

In *The Secret Joke of Kant's Soul*, Joshua Greene tries to show that this claim needs a qualification (Greene 2008; see also Singer 2005). He argues, partly on the basis of experimental evidence, that while projectivism is correct when it comes to Kant and other deontological philosophers, it is wrong in the case of consequentialists. This latter position is not based on automatic emotional responses, but on the epistemologically respectable capacity for explicit reasoning. (In a similar vein, Jonathan Haidt mentions the work of Peter Singer as a possible example of a non-intuitive approach in normative philosophy (Haidt 2001, 829).) Whether Greene is right about this depends on what we mean by consequentialism. If it is interpreted very modestly, as the thesis that it is, as seem from an impartial point of view, desirable that one performs the action with the best consequences for all the parties involved, then it may indeed resist projectivist debunking. However, interpreted in this way, consequentialism is not a complete claim about ultimate reasons. First, we are not told which consequences count as good and which count as bad. Greene takes it for granted that a consequentialist will say things like "better to save more lives" (op. cit., 39), but this is of course only so if it is better to live than to die. And why do we think that it is? An obvious suggestion would be that we rely on the intuitive capacity that Greene repudiates deontologists for putting their trust in. At any rate, he provides no alternative answer.

Second, consequentialism as defined above does not necessarily possess independent authority. If it is supposed to constitute a source of ultimate reasons, it must be the case that all individual agents somehow just have to care about what conduct is desirable from an impartial point of view. Again, it is hard to see how to justify such a claim other than by appealing to automatic and affectively laden intuitive judgments.

In this section, I have suggested that the belief in ultimate reasons could be a result of projection. I have also pointed to some empirical findings that seem to support this position. I

¹⁰³ Again, I am disregarding the neo-Kantian arguments criticized in earlier chapters.

do not want to claim that the findings, by themselves, demonstrate beyond any reasonable doubt that projectivism is correct. My aim has only been to show that there is a simple, straightforward and apparently plausible alternative to realism. And since realism is vulnerable to the objections from queerness and disagreement, we should definitely be willing to consider alternatives.

In the previous section, I pointed out how a realist needs to separate the intuitions that he considers fixed points and evidence for his position from more questionable intuitive convictions. The need for a systematic and principled distinction between good and bad intuitions is equally clear if we start from the opposite direction. Any realist will presumably acknowledge that the projectivist is right about many particular cases. Quite often, what appears to be an ultimate reason isn't really an ultimate reason at all. For instance, there are football fans who seem convinced that it is of intrinsic importance that their team does well. Or consider how young people these days (according to old people these days, at least) want to be famous just for the sake of being famous. Philosophers also go wrong, of course. In *Principia Ethica*, G. E. Moore tells us that "pleasures of lust" are intrinsically bad, since they involve "cognitions of organic sensations and perceptions of states of the body, of which the enjoyment is certainly an evil in itself". If we imagine a universe consisting only of a man taking pleasure in these sensations and perceptions, we "cannot avoid the conclusion that the existence of such a universe would be a far worse evil than the existence of none at all" (G. E. Moore 1903, 257-8). Joseph Raz makes the – to my mind, at least – extraordinary claim that "[p]laying tennis is intrinsically good. It can also be good instrumentally, as a way of keeping fit, making friends or money, or gaining prestige. But apart from any beneficial consequences playing tennis may or may not have it is a valuable activity; it is an activity with intrinsic value." (Raz 1999, 296)

A realist must hold that even if the projectivist, debunking account of normative judgments outlined above is correct for these particular cases, it is not correct in general. Claims about reasons occasionally, or perhaps even quite often, amount to nothing more than an emotional response generating a false judgment, but there is at least sometimes some sort of grasping of a normative truth going on. We must not overlook the huge normative, metaphysical and epistemological difference between these two types of cases. There really must be a dramatic distinction between when Moore is just venting his Victorian prejudices and when philosophers are correctly describing ultimate reasons. However, it seems, when considering all the various claims about ultimate reasons made by philosophers and others, far from obvious where this distinction could be located. The problem is not just that it is hard to

see exactly where the line should be drawn, in the same way that we might wonder where to draw the line between caution and cowardice. Rather, the worry is that wherever it is put, it will appear arbitrary and unable to carry its enormous normative, metaphysical and epistemological weight. Apparently, the factors that generate emotive-intuitive projections are present in all judgments about ultimate reasons. Judgments that are, to a bystander, clearly mere projections do not necessarily differ from other judgments when it comes to their introspective feel or the level of confidence with which they are made. Why not just accept the simple and straightforward view that all judgments about ultimate reasons are the result of projection?

These considerations constitute a serious challenge both for realism as a metaethical position and for any intuitively based first-order claim or theory that is intended to be taken in a realist sense. Realists often try to present their view as the default, common sense position. They think the natural assumption is that our reasons discourse does refer to something objectively real, and that the burden of proof is on anyone who wants to claim that it does not (Nagel 1986, 143-4; Brink 1989, chapter 2; Shafer-Landau 2003). I find this way of construing the debate questionable. It could with equal plausibility be said that projectivism is the default position, since we all agree that there are at least some cases where a debunking, projectivist account of normative judgments is correct.

8. The case for hedonic egoistic ultimate reasons

In this chapter, I suggest a solution to the problems presented in the previous chapter. I argue that there is a way of defending hedonistic egoism about ultimate reasons that allows us to give a systematic debunking explanation of the intuitions that conflict with this doctrine, without thereby putting the intuitions that support it into question. Moreover, there is no problem with extensive and persistent disagreement in the case of intuitions regarding the value of one's own phenomenal states. Also, while the notion of an experiential quality that one just has to seek is, undeniably, somewhat queer, it does not seem to be queerer than the alternative: there is no credible way of explaining away our immediate evaluations of our own phenomenal feels. If these claims are correct, they would serve to vindicate both realism about reasons for action and a distinctive first-order view on which reasons for action there are.

I begin by explaining how pleasure and pain play a crucial role in the projective mechanisms described in section 7.3. This means that while we can explain away non-hedonic intuitions in the way that projectivists have suggested, not all hedonic intuitions will be vulnerable to this form of criticism. In section 8.2, I discuss which normative hedonic judgments would survive hedonic debunking, arguing that only the ones that support hedonistic egoism would do so. My claims in these sections do not exclude the possibility that there are other forms of debunking, which could undermine even these judgments. I discuss whether there are in section 8.3.

8.1 Hedonic debunking of non-hedonic intuitions

As we saw in the last chapter, anti-realists typically rely on some form of projectivism when explaining why it often seems as if there are ultimate reasons. However, the nature of these subjective responses is usually not spelled out in any detail. This is, I believe, a critical oversight. More specifically, would-be projectivists need to take into account the crucial role hedonic reactions appear to play in the formation of evaluative beliefs. There are good reasons to think that non-hedonistic and non-egoistic normative intuitions can all be explained away as results of hedonic projection. In this section, I will sketch how an explanation of this kind would go.¹⁰⁴

¹⁰⁴ A similar explanation is suggested in Brax 2009, 2.5.4. See note XXX.

In section 7.3, I outlined how our judgments about reasons tend to stem not from explicit reasoning, but from a fast, automatic and unconscious information-processing system. Hedonic processes are an integrated part of this cognitive capacity; the system is, as Seymour Epstein puts it in a pioneering paper, “pleasure-pain oriented” (Epstein 1994, 711; see also Casebeer 2008; Hynes 2008). Paul Slovic has dubbed this phenomenon “the affect heuristic”:

Representations of objects and events in people’s minds are tagged to varying degrees with affect. In the process of making a judgment or decision, people consult or refer to an “affect pool” containing all the positive and negative tags consciously or unconsciously associated with the representations. [...] Using an overall, readily available affective impression can be far easier—more efficient—than weighing the pros and cons or retrieving from memory many relevant examples, especially when the required judgment or decision is complex or mental resources are limited. (Slovic 2002, 1335-6)¹⁰⁵

As is the case with many other heuristics discovered by psychologists, we do not in general notice that we apply this strategy. It is not as if we explicitly tell ourselves “this is a complicated judgment and I don’t really have the time to think things through, so I’ll just look at some representations in my head and notice how I feel about them.” Also, there is no explicit or conscious *inference* from one’s hedonic level to an evaluation of a non-hedonic object. Rather, the negativity or positivity is automatically projected: it is experienced as belonging to whatever else is going on in our minds. As Gerald L. Clore and Jeffrey R. Huntsinger put it, “affect assigns value to whatever seems to be causing it.” (Clore & Huntsinger 2007, 393)

A classical study by Norbert Schwarz and Gerald L. Clore provides a nice example of this phenomenon (Schwarz & Clore 1983). The participants were all asked to rate their satisfaction with their lives as a whole. Some were interviewed on a rainy day, others on a sunny day. It turned out that the first group was substantially less satisfied. The explanation provided by Schwarz and Clore is that these participants were in a bad mood because of the weather and then took their low hedonic level to reflect a lack of success in their lives. When they were explicitly asked what the weather was like that day, their ratings were almost as

¹⁰⁵ It is not always clear what psychologists mean by “affect”. Slovic defines it, in the abstract of the paper, as “the specific quality of “goodness” or “badness” (i) experienced as a feeling state (with or without consciousness) and (ii) demarcating a positive or negative quality of a stimulus.” Presumably, (ii) is derived from (i): a stimulus is positive to the extent that it provokes a good feeling state. It would seem, then, that Slovic’s conception of affect is very similar to the conception of hedonic states I defended in chapter 3.

high as those in the other group – presumably because they then attributed most of their lack of positive affect to the bad weather instead of their life in general.

In other words, we tend to value those things that we are doing or thinking about when we feel good. We generally enjoy discovering what the world is like, and so we end up according intrinsic value to knowledge. Since helping others makes us feel good and seeing them suffer makes us feel bad, we come to believe that their well-being is a source of ultimate reasons to us. And so on. If we look back at the list of proposed intrinsic values presented in the last chapter, every single one of them is in some way related to positive affect.

It is important to stress that the outcome of this process is not (only) *desires*; it is not just that we have a glowing feeling when thinking about, say, knowledge, or that we feel some urge to seek it. The hedonic mechanisms generate intuitive *judgments*, and these judgments aren't in any way about pleasure. They have the simple form "X is good", and not "X is a source of pleasure", "X is good as evidenced by my hedonic reaction" or anything like that. It is the cause of the intuitions that is hedonic, not the content. Mill was mistaken when he thought that people value virtue and other non-hedonic goods "as a part of their happiness" (Mill 1871, ch. 4, par. 5). Also, when we act on these judgments, there need not be any ulterior hedonic motive. We may seek knowledge or power or freedom for its own sake – it is just that in the explanation of why we have this non-hedonic motivation, hedonic states play a crucial part. Thus, in order to form non-hedonic motives and evaluations, we do not first, as Mill seems to suggest (*ibid.*, ch. 4), have to go through a stage where the object is sought or valued merely as a means to pleasure.

Our normative intuitive responses seem, then, to be shaped by hedonic experiences. Often, the resulting responses themselves are also affectively laden. When we think of, say, hanging an innocent man, we feel bad about it: it is an unpleasant scenario. However, intuitive responses need not always have this character. As Seymour Epstein says, "Affect is assumed to play an important role in the acquisition of information in the [automatic] system, but as behavior (including mental behavior) is practiced, it becomes increasingly proceduralized and affect free." (Epstein 1994, 714) For instance, a philosopher who intuits that knowledge is an intrinsic value need not feel any discernable pleasure when he thinks about knowledge, especially if his intuitive judgments was formed a long time ago and has since been frequently rehashed. It may present itself as just as unbiased and emotionally neutral as any other philosophical intuition. In other words, the hedonic origin of normative intuitions is very difficult to uncover through introspection, and this would explain why so many philosophers could have been lead astray on this count.

If our judgments about reasons really have a hedonic origin, one might expect us to accord intrinsic value to various objects and activities roughly in proportion to their instrumental value as sources of pleasure. Sidgwick thought that we do in fact do so, and used this correlation in his argument for hedonism (Sidgwick 1907, 401-2). Though there is presumably an element of truth in his observation, the full picture is more complicated. There are a number of reasons why one might end up valuing something out of proportion to its hedonic merits – even if we assume that the evaluation is generated solely by affective projection.

First, an object may become liked or disliked because it is associated with another object or activity that is hedonically positive or negative. Psychologists refer to this phenomenon as evaluative conditioning (for an overview, see de Houwer, Thomas & Bayens 2001). To provide a somewhat trivial illustration, participants in an experiment by Jones, Fazio and Olson were briefly shown a picture of a neutrally looking cartoon character simultaneously with an affectively laden word or image during what they took to be a study on attention. When they were later asked to evaluate the character, those who had seen it presented together with a positive word or image liked it significantly better (Jones, Fazio & Olson 2009).

Evaluative conditioning presumably plays an important role in socialization. When children break a promise, for instance, they are met with parental blame or punishment, which makes them feel bad. Later, the thought of breaking a promise will provoke a negative hedonic reaction, and so children come to believe that there is something inherently bad about this kind of conduct. Eventually, the mere knowledge that your parents (or you peers, or someone you admire) judge something to be good or bad may lead you to feel good or bad about it yourself, and consequently to share in their judgment. Thus, a complete hedonic projectivist explanation for why a given person values a given object is likely to be very complicated.

Evaluative conditioning takes place more easily if the subject does not realize what is going on. The participants in the experiment mentioned above were not aware of the fact that they had previously seen the cartoon character in conjunction with a positive word or image. Accordingly, they automatically took their positive reaction as evidence of there being something attractive about the cartoon character itself. If they had been aware, they would presumably have attributed some or all of the reaction to the word or image, and so showed less liking for the character. Similarly, if children were to attribute all their negative feelings concerning promise-breaking to previous or imagined negative reactions from their parents,

they would not, or at least not to the same extent, come to think of promise-breaking as bad in itself, but merely as something their parents don't tolerate. Evaluative conditioning is thus facilitated by hedonic misattribution (Jones, Fazio & Olson 2009).

It turns out, then, that we form likes and dislikes to objects we are thinking about when we feel good or bad, even if there is no direct connection between the object and the hedonic reaction. In fact, objects may come to be valued through evaluative conditioning even if the subject has never had any hedonically positive experiences when encountering the object at all. This happens when two objects become associated without any hedonic reaction or evaluation taking place and one of the objects later becomes the target of evaluative conditioning. In these cases, the liking will to some extent generalize to the second object as well. This phenomenon is known as sensory preconditioning, and occurs in both classical and evaluative conditioning (see de Houwer, Thomas & Bayens, 862). To illustrate, a person may associate fame with power without valuing either of these things. If he subsequently develops a liking for power, he may also come to like fame for its own sake – even if we assume that he never enjoyed fame or had any positive hedonic reactions when thinking about it.

A further complication is the fact that once an attitude towards an object is formed through evaluative conditioning, it can remain in place for a long time even if further experiences with the object are hedonically neutral. Extinction is very slow in evaluative conditioning compared to other forms of conditioning (see de Houwer, Thomas & Bayens, 858-9). A philosopher who originally became attracted to knowledge because of its hedonic benefits, or through a more elaborate route of evaluative conditioning, may persist in his intuitive valuation of knowledge even if his later experiences with seeking or gaining knowledge are not particularly hedonically attractive.

The account given here is meant to apply both to ordinary people and to non-hedonist and non-egoist philosophers. Now there is no reason to deny that these philosophers often override their intuitive responses, or that their coherentist reasoning is generally perfectly respectable. It is not as if all books in normative philosophy are simple reflections of the hedonic experiences of their author. My suggestion is more modest: in the justification of any complete non-hedonic or non-egoistic reason claim, one always finds, at some crucial stage, appeals to intuitions, and these intuitions are most plausibly seen as hedonic projections, generated in the way I have just outlined.

We do not presently have sufficient empirical evidence to conclude that this hedonic explanation of non-hedonic valuing is correct. Many of the experiments do not involve full-fledged judgments about ultimate reasons, and none of them have professional philosophers

as participants. The picture I have sketched may need substantial revisions, and it is far from complete. There is reason to think that, once hedonic projection has got the formation of evaluative beliefs going, we may be prone to further errors and biases, at least in everyday life. As an example, people have been shown to make many decisions based on a heuristic of “anchoring and adjustment”: we start out with some sort of position that we take to be the default one, and then proceed to make adjustments (which are usually too small) in the light of the evidence we are presented with. In experiments, researchers have shown that the placement of the anchor can sometimes be completely arbitrary. Thus, participants who had just been told to write down the last two digits of their social security number subsequently turned out to be willing to pay a sum in the same range for a bottle of red wine (Ariely 2008, ch. 2). I see no reason to think that hedonic states play any direct role in generating this bias. The claim I want to make is that pleasure and pain supply the basic positivity and negativity needed for us to take a normative interest in our surroundings, and not that the details of every single value or reasons judgment is fixed exclusively by previous hedonic experiences.

Even if we need, then, to await further research before drawing any final conclusions, I do think that it is at least safe to say that attempting to uncover a hedonic explanation of non-hedonic reasons judgments is a more promising research programme than doing non-hedonistic or non-egoistic normative philosophy. While the latter project has to deal with all the problems discussed in the last chapter, I see no obvious reason why the former should not turn out to be successful.

If the form of hedonic projectivism outlined here is on the right track, it should come as no surprise that there is so much disagreement in normative philosophy. In particular, the hedonist can explain why almost all people have failed to see that his position is correct, and, importantly, he can do so without undermining the epistemological credibility of the normative intuitions that underlie this position. If other people reject hedonism because they make hedonic projective errors, then this should not reduce our confidence in the intuition that there are hedonic reasons.

It is clear that no other monistic theory of reasons could give such an account. For instance, it would be wildly implausible to claim that we value pleasure, freedom, friendship and power only or mainly because they have somehow been associated with knowledge. Of course, a pluralist might try to show that some people value the things that are not on his list of values only because they have been associated with things that are. I find it unlikely that any such explanation would succeed. At least it is clear that there aren't presently any specific explanations of this kind on the table. As we saw in section 7.2, the explanations realists have

given for why there is so much disagreement in moral philosophy are all *general*, and so lead to a correspondingly general skepticism about the justificatory powers of normative intuitions.

8.2 Which intuitions survive hedonic debunking?

I have claimed that the existence of hedonic projection should not undermine our confidence in all intuitive judgments concerning the value of pleasure. We do not give a complete and satisfactory explanation of why we value pleasure by saying that we react with pleasure to pleasure and that the pleasure is then projected onto the pleasure. I have, however, left it open precisely which intuitions concerning pleasure and pain that are immune to hedonic debunking. In this section, I examine this issue. I will argue that the line is most plausibly drawn at the intuitions that support hedonistic egoism as a theory of ultimate reasons.

Our most basic hedonic evaluations are those that are immediate and concern only the quality of our own present mental states. These evaluations are needed to get the whole process of hedonic projection going in the first place and could not possibly be explained away as a result of this process (though there are, as we will see in the next section, other ways of questioning them). As Leonard Katz notes, “it is hard to see how the seeming perception of one’s own pleasure and of its goodness or of one’s own suffering and of its badness could be discredited by being shown to be based in affective reactions.” (Katz 2008, 410)

It is not obvious what the content of these primitive hedonic intuitions are. First, there is the question of whether they are personal in nature. It could be claimed that, whenever a person experiences pleasure, he directly apprehends that this is (1) intrinsically good for all agents, (2) intrinsically good for himself, or (3) just intrinsically good, in a primitive sense that is not personally determinate. I find (1) implausible, and I am not aware of any authors who have defended it.¹⁰⁶ If (2) is correct, then we need no further argument or intuition to establish the existence of egoistic hedonic value. If (3) is correct, on the other hand, we would have to make an additional claim that if an experience is in some way intrinsically good, then the goodness must at least concern the person whose experience it is. I can think of no reason – hedonic-projective or otherwise – to question this claim. (It would be a further question, to which I return below, which normative status the experience has for other agents.)

¹⁰⁶ Thomas Nagel claims that the way pain is subjectively experienced gives us reason to think that there are agent-neutral hedonic reasons (Nagel 1986, VIII.5). However, his idea is apparently not that we directly apprehend (1), but that (1) is supported by (3). I discuss and criticize his argument in section 9.3

Second, there is the question of whether the fact that one's present phenomenal state is good or bad automatically entails that there is any reason to *do* anything. It is far from obvious that it does. After all, the present state itself cannot be improved, since one cannot affect the present, but only the future. Thus, it seems possible to be a sceptic about hedonic reasons without being a sceptic about hedonic value. Some philosophers hold a view that would make this form of scepticism impossible: they think that claims about value can be reduced to or equated with claims about reasons for action. According to Scanlon's so-called buck-passing theory of value, for instance, being valuable is not a property in its own right, but only the "purely formal, higher-order property" of having lower-order properties that ground reasons for action (Scanlon 1998, 95-100; see also Parfit forthcoming, 72-6). However, it seems to me that what is misconceived here might be the present way of analyzing value, rather than the present form of scepticism. Scanlon never considers this issue, and we should not automatically assume that because his view has generally plausible implications if we disregard worries like this, it must also have the right implication about this particular matter. Moreover, the buck-passing account of value is also controversial for other reasons (see J. Olson 2004; Crisp 2005). We cannot, therefore, disregard the possibility that there may not be any hedonic ultimate reasons even if there is such a thing as hedonic value.

In section 6.1, I described how one might be led to this position if one accepts a radical reductionism about personal identity: there is no reason to try to improve one's hedonic level, since by the time the improvement occurs one would no longer be the same person. The present worry is even more radical: even if we disregard worries about personal identity, it may not be the case that one's present self is normatively bound by what will happen to one's future self.

It is important not to confuse this view with another with a form of egoism that might have a certain (very moderate) intuitive appeal. In *Reasons and Persons*, Parfit describes a theory that he calls hedonistic egoism about the present and says that it has the "absurd" implication that one should end a presently occurring pain by pushing a button even if the consequence is that, after a few minutes have passed, one's life will become terrible (Parfit 1984, 134). This view is, however, not egoism of the present, but egoism of the very, very near future. Even pushing a button takes some time. Egoism about the present implies that there isn't even a reason to push a button to end your pain even if doing so will have no bad long-term consequences. All the view allows us to say is that suffering is bad for you at the time it is experienced.

Most people undoubtedly have a strong intuition that they do have a reason to do something about their hedonic levels. But it could be claimed that this intuition is the result of projection: the idea of future suffering makes you feel bad now, and so you project the importance of the present hedonic state onto the future hedonic state. Of course, it wouldn't be a mistake to think that your future pain would matter to you *then*, but it might be a mistake to think that it matters to you now.

Ideally, I would like to be able to show how this reasoning could not possibly be correct. Sadly, I see no way of doing so. I will have to be content with just stating that the view seems crazy and obviously false. The problem is not that I think it is out of the question to accept its implications; we might, as I have said, be led towards the same first-order view if a radical form of reductionism about personal identity should turn out to be correct. It is, rather, that I find it difficult simply to get my head around the view itself. If we assume that the Ivar Labukt who exists a second from now is exactly the same person as the present version, and that his hedonic level will matter to him then, how could it possibly fail to matter to me now? I have the impression that this is hard to accept for cognitive and not hedonic reasons. Of course it is conceivable that I am misled and that the present form of hedonic projection just goes especially deep. I do, however, think that there is at least a decent chance that I am right.

If we assume that there is a reason to be concerned with one's future hedonic states, a further question is whether we should be neutral between different times. One possibility is that there is a certain point in the future where the scope of our present hedonic reasons suddenly ends. This view seems unacceptably arbitrary no matter where the limit is drawn. It is hard to believe that one's future pains would start being important, say, 3,2 seconds or 54 days before they might be experienced. For this reason, the idea of a discount rate for future pleasures and pains might seem more promising. Again, we are disregarding the issues concerning personal identity mentioned in chapter 6, so the idea would be that temporal distance itself is normatively relevant. I doubt this, as do most authors.¹⁰⁷ I find it implausible to take this judgment as a result of hedonic projection. It is, after all, a purely negative judgment, and it does not seem likely that virtually all philosophers have had so positive hedonic experiences with strict temporal neutrality that they come to mistakenly deny that

¹⁰⁷ This is not to say that the question "when?" cannot have any normative significance. As Parfit points out, we do not think that our own past pains matter directly to us (Parfit 1984, sect. 64). (If we accept temporal neutrality concerning the future because of a prior general conviction that time is normatively irrelevant, this might be a problem. However, as Parfit notes (*ibid.*, 179-81), one need accept it on that ground, but simply claim, as I have done here, that it is independently plausible.)

temporal distance really does matter. It would be much more plausible to claim that the bias towards the near future displayed by a significant number of people (in their actions, if not in their judgments) could be explained by the fact that the prospect of immediate pleasure or pain provokes particularly strong hedonic reactions.

I turn now from the normative status of one's own future hedonic states to the normative status of the hedonic states of other people. While we cannot directly evaluate the quality of other people's phenomenal states, the reasonable assumption is surely that they experience something very similar to what we do. As explained in section 5.1, even the egoist can accept that any agent has reasons to seek his own pleasure and avoid his own pain. The present question is whether there is also an ultimate reason to be concerned with hedonic states that one does not get to experience.¹⁰⁸

There is no doubt that most people, philosophers included, have an intuition to this effect. It also seems clear that a hedonic projection could at least to a great extent account for this fact. The idea of other people suffering tends to provide a direct negative hedonic response and is associated with many other things we disvalue, while the idea of other people enjoying themselves has the opposite effect on us. It might be objected that the non-egoistic hedonic intuitions are more universally shared and typically also stronger than all the other intuitions that the mechanisms described in the previous section are supposed to explain away, and so even if the idea of giving hedonic debunking explanations of normative intuitions is generally plausible, it should not automatically be assumed that it works in the present case. As explained in section 7.2, this would only be a good objection if it would be difficult for the would-be debunker to explain why the intuitions in question are so prevalent. Apparently, it is not. First, our affective reactions to the suffering and enjoyment of other people are often very strong. These sympathetic reactions occur from a very early point in life, and do not seem entirely attributable to any form of conditioning, so they seem to be at least to some extent innate. There is also evidence suggesting an innate disposition for developing automatic affective judgments concerning how it is permissible to treat other people; across cultures, children do so in similar ways and at a similar age (see Joyce, 4.5). Second, affective reactions and corresponding intuitive judgments concerning how one treats or is treated by others are of course socially reinforced – to a greater extent, I would think, than in any other

¹⁰⁸ If one thinks that there are no hedonic reasons but only hedonic value, one could ask the analogous question about value. I believe the answer would be analogous as well.

case. For these reasons, the mere fact that a large majority of people have some very firm non-egoistic intuitive hedonic convictions should not make us think that these intuitions are especially resistant to hedonic debunking.

The interesting question is whether we have, in addition to what we would expect hedonic projection to cause, some more cognitive and unassailable intuition that there are agent-neutral hedonic reasons. In *The Methods of Ethics*, Sidgwick at one point seems to claim just that. More precisely, the intuitions that one ought to give no special priority to oneself or to any particular time when seeking pleasure and avoiding pain is supposed to be analogous to or have some affinity with the “abstract” intuitions we have regarding “the individuals that make up a Logical Whole or Genus” and “the similar parts of a Mathematical or Quantitative Whole” (Sidgwick 1907, 380-3). (Sidgwick also had an intuition that egoism is correct (see e.g. *ibid.*, 497-8), but we may disregard that for the moment.) It is difficult to understand how an egoist would be making a mistake that is similar to any mathematical or logical mistake. Of course, this could be the case if he started out by accepting that all pleasure is good in the agent-neutral reason-implying sense, so that there is no normatively relevant difference between his own pleasure and that of others, and then just arbitrarily decided to disregard the pleasure of everyone but himself. But hedonistic egoism is the doctrine that there *is* a normatively relevant difference between one’s own pleasure and that of others, namely that one only gets to experience the former. Whether this difference really is normatively relevant is a question that doesn’t seem to have anything to do with logics or mathematics. (Similarly, we might have an “abstract” intuition that if it is irrelevant how far into the future a pleasure is, then it is a mistake to prefer a smaller pleasure today to a greater pleasure tomorrow. But whether temporal distance matters in the first place cannot be settled by the same kind of intuition.)

Some philosophers might claim that they have some sort of intuition in favor of agent-neutral hedonic reasons that is, even if not on a par with mathematical intuitions, then nevertheless more cognitive in nature than those that stem from hedonic projection, and so potentially more epistemologically respectable. It is difficult to argue with a claim like this. For what it’s worth, I find that I have exactly the same kind of intuitive propensity to believe in non-egoistic ultimate reasons as I have to believe in non-hedonistic ultimate reasons. There seems to be a reason to please others and to alleviate their pain in the same way as there seems to be a reason to keep promises, seek knowledge or refrain from hooking up to an experience machine. To be sure, some of the intuitive non-egoistic appearances, especially those concerning children, kind people or people I know, are very vivid – perhaps more so

than any of the non-hedonistic appearances. Still, this seems to be a matter of degree rather than kind. After considering how my non-egoistic convictions might be just hedonic projections, I have the impression that they are precisely that. There are no further, purely cognitive non-egoistic intuitions to be found. When I try to rise above or adjust for affective influences on my reasoning, it seems to make all the difference in the world whether a given pleasure or pain will be had by me or by someone else. I don't experience what other people experience, and so I don't have to care about it – it's that simple. In fact, from this perspective, my prior non-egoistic beliefs seem almost incomprehensible. Perhaps the best way of putting it is to say that while I can understand *why* I thought that I just had to respect the hedonic interests of other people, I cannot understand *how* I could think so.

My defense of the egoistic version of hedonism would of course be strengthened if I could point to some obvious difference or disanalogy between intuitions regarding personal and temporal neutrality that would show how the debunking explanation is correct in one case and out of the question in the latter. I am not aware of such a difference; as I have said, I just have different considered judgments regarding these cases. Some philosophers apparently find it implausible that there should be such a difference. In *Reasons and Persons*, Parfit observes that rational egoism, or S, is a “hybrid theory”, since it “rejects the requirement of personal neutrality, but requires temporal neutrality.” He then goes on to claim that

as a hybrid theory, S can be charged with a kind of inconsistency. If an agent has a special status, why deny this status to the time of acting? [...] This [charge] can appeal to the analogy between oneself and the present, or what is referred to by the words ‘I’ and ‘now’. This analogy holds only at a formal level. Particular times do not resemble particular people. But the word ‘I’ refers to a particular person *in the same way in which* the word ‘now’ refers to a particular time. And when each of us is deciding what to do, he is asking, ‘What should I do *now*?’ Given the analogy between ‘I’ and ‘now’, a theory ought to give both the same treatment. (Parfit 1984, 140)

Parfit appears to consider this an independent and serious objection to any version of rational egoism, including the hedonistic one.¹⁰⁹ I find it hard to understand what the objection would be. Why should there be a requirement that a normative outlook must have a general policy on all issues that can be framed in terms of neutrality and relativity? In order to see how unreasonable such a requirement would be, imagine a person who believes in hedonistic

¹⁰⁹ On a preferentialist understanding of self-interest, the egoist faces some additional complications (see *ibid.*, sect. 59-61). Thus, it may be that Parfit's argument has some force against this kind of egoism.

egoism of the present. At a particular moment, he judges that his present mental state is bad for him. You then ask him whether it would have been better or worse for him if he had had the same experience in a different place. The only reasonable answer is, of course, that one's location is irrelevant to the value of one's experiences. However, by Parfit's reasoning, the person could not make this reply, since he would then be holding a "hybrid" theory which could be charged with "a kind of inconsistency", since it involves full personal and temporal relativity in conjunction with full spatial neutrality.

It might seem uncharitable to interpret Parfit as endorsing a completely general requirement on a fixed neutrality policy. Perhaps he only means that a plausible theory should be equally neutral when it comes to times and persons. However, Parfit explicitly says that the analogy "only holds at a formal level" and that "particular times do not resemble particular people", so his point couldn't be that there is some especially important analogy between times and persons.¹¹⁰ His view would have to be that a theory of reasons must either be "for" or "against" neutrality. This is just silly. Each of the issues that can be described in these terms should be decided on its own merits. If one can show that it is plausible to say that it is irrelevant when a future pleasure will be experienced, and if one can also show that it is plausible to say that it matters greatly whether the person experiencing it will be you, there is no *further* task of showing how both these things could be true at the same time.

Of course, it could happen that when defending the first claim, one uses a form of argument that would, if valid, undermine the second claim. Parfit mentions one way in which this might be the case: if we think the observation that the temporal position of a pleasant experience doesn't affect its qualities constitutes a sufficient defense of temporal neutrality, we would have to accept that a similar argument can be made regarding personal neutrality, which again would force us to admit that egoism is mistaken (*ibid.*, sect. 63). However, it is by no means necessary for egoists to proceed in this manner, and I have not done so in this section. My claim has been that the conviction that one ought to promote one's future hedonic level in a temporally neutral way is more cognitive in nature, and less obviously a candidate for hedonic debunking, than the conviction that one ought to promote the hedonic levels of other people. One may, of course, deny that there is such a difference. But then the problem

¹¹⁰ In fact, Parfit does believe that similar considerations apply to spatial positions, but he takes this to be practically irrelevant, since a person can only be in one place when deciding what to do (*ibid.*, 140-1). This is true, but we can still ask hypothetical questions, and the answers to these questions shouldn't be determined by one's attitude towards the normative importance of the distinction between persons.

would be that my defense of egoism rests on a false claim, and not the mere fact that egoism is a “hybrid” theory.

Finally, there is the issue of the relative values of different forms of pleasure. As noted in section 4.1, if the pluralistic hedonic tone view is correct, some positive tones could be more valuable than others even if their intensity is the same. If this is what is meant by saying that some pleasures are inherently better than others, the claim could be based on immediate evaluations of actual experiences, and nothing I have said in this chapter would tend to undermine it.

One can also claim that certain pleasures can be better than others because of their causes or objects. Now the different positive hedonic tones may of course have different characteristic sources; one of them might, say, be primarily associated with bodily sensations and another with intellectual pursuits. If so, we could in a loose sense say that the value of a pleasure varies with its object. Strictly speaking, though, the object would not be relevant for the goodness of the experience; what matters is only the phenomenal quality of the various hedonic tones. Many philosophers have, however, thought that the object can affect the value of a pleasure independently of its influence on how the pleasure feels. When Moore denounced (at least certain forms of) sexual pleasure, his objection was not that it doesn’t feel good enough, but that it involves taking pleasure in inappropriate sensations. Mill is sometimes interpreted along the same lines by his critics (see e.g. 1907, 94-5).

More recently, Fred Feldman has suggested a view called “desert-adjusted intrinsic attitudinal hedonism”, which explicitly lets the value of hedonic experiences depend on their objects. He thinks that

it is reasonable to describe certain affairs by saying that they ‘deserve to be the objects of pleasure’. In the case of such objects it is fitting, or appropriate, that someone takes pleasure in them. Thus, for example, consider the state of affairs that consists in some painting being genuinely beautiful.¹¹¹ It is reasonable to say that this state of affairs deserves to be appreciated. [...]he intrinsic value of an attitudinal pleasure is determined not simply by the intensity and duration of that pleasure, but by these in combination with the extent to which the object of that pleasure deserves to have pleasure taken in it. (Feldman 2004, 119-20)

¹¹¹ It is far from clear that there is such a thing as a picture being “genuinely” beautiful, as opposed to merely tending to provoke a certain hedonic reaction from human beings in general or from some more limited audience. Let us, however, disregard this complication.

Feldman justifies this view by appeals to intuition. “If”, he says, “we tweak our hedonism by adding some assumptions about the pleasure-worthiness and pain-worthiness of certain objects, we can get the theory to yield evaluations consistent with our firm and unshakable pre-analytic assessments.” (ibid., 122) There is a worry here that the assessments in question are ultimately based on intuitions in favor of pluralism: the enjoyment of a beautiful painting seems to be especially valuable because beauty itself seems to be valuable.¹¹² If so, and if the pre-theoretical convictions really are firm and unshakable, it would be more reasonable to accept some form of value pluralism, which is, when compared with the desert-adjusted version of hedonism, a more straightforward and less paradoxical view.¹¹³

More importantly, even if we assume that Feldman’s pre-analytic assessments aren’t just pluralistic intuitions in disguise, it is still likely that they are results of hedonic projection. Usually, the pleasures philosophers take to be particularly valuable are the pleasures of a highly educated intellectual. The pleasures that are supposed to be inferior are typically those that are derived from the misfortunes of others, display a lack of human perfection, or have bad long-term consequences. In all these cases, providing a hedonic debunking explanation seems to be quite straightforward.

I believe, then, that only intuitions in favor of agent-relative reasons to seek and avoid certain phenomenal qualities would survive hedonic debunking. The reader may of course think I am wrong about this. It is important to keep in mind that this would be a disagreement not over whether my general argument in this thesis works, but over what form of hedonism it supports.

8.3 Can basic hedonic value judgments be debunked?

¹¹² This worry would not apply to the common intuition that sadistic pleasure is inferior or even positively bad (see e.g., G. E. Moore 1903, 262-5; Zimmermann 1980; Brink 1989, 263-4). (Ross makes the related claim that only deserved pleasure is good and undeserved pain is bad (Ross 1930, 136-9), and what makes a person deserving could be cashed out in purely hedonistic terms: you are deserving to the extent that you have provided others (and yourself, perhaps) with pleasure.) This view sits ill with the egoistic interpretation of hedonism, however. It is implausible to hold that while there is no ultimate reason to prevent others from feeling pain, there is an ultimate reason for avoiding pleasures that have the pain of others as their object or cause.

¹¹³ In fact, Feldman only says that pleasure is the only thing that makes a *life* intrinsically good, i.e. that it is the only component of welfare (ibid., chapter 1). He could consistently claim that other things, such as beauty or knowledge, are intrinsically valuable even if they do not by themselves make anyone’s life go better. Mill would have rejected this possibility; Feldman does not consider it.

As I have said, the fact that claims about the value of one's own phenomenal states are not undermined by having a hedonic origin does not mean that they cannot be debunked in some other way. Whether they can is the topic of this section.

The simplest way of resolving this question would be to insist that our basic hedonic judgments just couldn't possibly be mistaken. According to Thomas Nagel,

we cannot from an objective standpoint withhold a certain kind of endorsement of the most direct and immediate subjective value judgments we make concerning the contents of our own consciousness. We regard ourselves as too close to those things to be mistaken in our immediate, nonideological evaluative impressions. No objective view we can attain could possibly overrule our subjective authority in such cases. There can be no reason to reject the appearances here. (Nagel 1986, 158)

It certainly wouldn't diminish the plausibility of hedonistic egoism if Nagel is right on this point¹¹⁴, but I am not sure that he is. It seems somewhat dogmatic to dismiss the possibility of systematic error in our hedonic evaluations from the outset. As regards their initial epistemological credibility, they should probably be considered appearances of value on a par with any other appearance of value. They, too, could be overturned by a credible debunking account of how they are generated. It is just that, as it turns out, such an account is hard to come by. (Or so I will argue.) Conceivably, after we have failed to undermine the hedonic intuition, we might end up thinking that human beings do not in fact ever make mistakes when judging the value of their experiences. For all I know, this could be true. But the important point is that the reliability of judgments about hedonic value must still be verified by coherentist reasoning, and not established simply by having intuitions that one immediately knows could never be mistaken.

We need, then, to examine possible reasons for thinking that that basic hedonic value judgments aren't correct in the realist sense. The first thing to note is that there is no argument from disagreement to worry about within this class of intuitions. On a first-order level, and when it comes to immediate evaluations of actual experiences, virtually everyone is for pleasure and against pain. We may not have any guarantee that this really is a sign of agreement: perhaps the phenomenal feels that we call pleasant differ from person to person. But in the present context, it is sufficient to note that there is at least no evidence of extensive

¹¹⁴ Nagel goes on to argue that the subjective appearances actually suggest that there are agent-neutral hedonic reasons. If this claim is right, the plausibility of egoism would certainly be undermined. I criticize his argument in section 9.3.

disagreement: we have no positive reason to think that human beings often make conflicting judgments about one and the same hedonic tone (though they do of course have different hedonic reactions to one and the same non-hedonic phenomenal quality, as explained in section 3.2).

In section 7.1, we saw that a realist outlook on reasons or values can also be accused of being queer and mysterious. In order to assess whether this objection applies in the case of hedonic value, I must first say a few words about what kind of thing hedonic value is supposed to be.

I have argued that pleasures and pains involve certain phenomenal qualities, or hedonic tones. It is in these tones that the intrinsic goodness or badness of hedonic states would have to reside. But how should we understand the notion of residence? One possibility is that positive hedonic tones have, in addition to their phenomenal properties, a non-analyzable property called “being intrinsically good”. This is the outlook on intrinsic value favored by G. E. Moore (G. E. Moore 1903, ch. 1). If we accept Scanlon’s buck-passing account of value, on the other hand, pleasant experiences are valuable because they have the second-order property of having phenomenal qualities that there is a reason to seek.

On both these accounts, there is a clear distinction between descriptive and normative properties. One can give a complete factual description of any entity with intrinsic value, including pleasure, without committing oneself to any normative claim. As Scanlon says, the fact “that a thing is pleasant” and other “natural” facts, may “provide the grounds” for conclusions about reasons for action. However, “judging that these facts obtain need not involve explicitly drawing these conclusions.” (Scanlon 1998, 96) In a similar vein, C. I. Lewis suggests that we first “describe the given [experiential] content in *other* respects than value” and then “make shift to evaluate it from that description” (C. I. Lewis 1946, 402). The idea is that we can talk or at least think about what it is like to experience pleasure and pain while still remaining agnostic about whether it is intrinsically good or bad to be in this state.

This way of looking at the relationship between fact and value appears plausible when it comes to non-hedonic candidates for intrinsic value. For instance, it would seem as if those people who think that promise-breaking is intrinsically bad and those who don’t can agree on what promise-breaking *is*, i.e. on how it should be described. What they disagree on is whether the thing meeting this description is an independent source of ultimate reasons. In the case of hedonic tones, however, the distinction between fact and value seems much more difficult to draw. I, for one, find it doubtful whether one can talk about or even imagine the phenomenal quality of pain – what it is like to suffer – stripped of all its normative properties.

As Joseph Mendola says, “the *phenomenal* difference between pain and pleasure seems to be at least in part that the phenomenal component of the former is nastier, intrinsically *worse*, than that of the second. [...] No one, not even a Martian, could give a complete and adequate characterization of that phenomenal state without talking about its nastiness, without making a committing mention of its intrinsic disvalue”. (Mendola 1990, 702) In other words, the badness or normative authority of pain may not be something that comes in addition to or is grounded in the phenomenal quality. It seems, rather, that the badness somehow just *is* the phenomenal quality. David Brax endorses the same view when he says that pleasure is a direct and immediate “experience of value” (Brax 2009, 218-20). Korsgaard tells us that “the utilitarian [she presumably means the hedonist] claims that pleasures and pains are facts which are also values, a place where the natural and normative are one.” (Korsgaard 1986, 146) I find this account quite plausible, but my defence of hedonistic egoism is also compatible with the alternative views.¹¹⁵

If we want to defend the claim that there are hedonic egoistic ultimate reasons, we are committed to holding that there is some sort of epistemological route to these reasons. The reason we believe that pleasure is valuable must be that it really is valuable. Several authors have thought that we have the ability to intuitively grasp the value of our hedonic states (see e.g. Tännsjö 1998, 117-21; S. Rachels 2003, sect. 4). This could either mean that we automatically infer that there is a reason to seek or avoid positive and negative hedonic tone or, if Mendola’s view is right, that we simply know what we experience. We need not in any case assume that the hedonic intuition is indubitable.

I have to admit that I see no way of explaining away the queerness of hedonic ultimate reasons: I really think pleasures and pains are, in Mackie’s words, “utterly different from anything else in the universe” (Mackie 1977, 38). But could it at least be argued that the intrinsic (dis)value of one’s own phenomenal feels would be less queer than non-egoistic or non-hedonic forms of (dis)value?

It might perhaps seem as if the account of hedonic value defended by Mendola could be of some help here. As we have seen, part of what made Mackie suspicious of ultimate reasons is the supervenience relation that is supposed to hold between natural and normative properties. If there is no such relation – if the badness is in some sense indistinguishable from the

¹¹⁵ If there is no clear distinction between the normative and factual properties of hedonic states, a hedonist could not formulate his criterion of what one ought to do in purely empirical terms. This might perhaps give us a reason to hope that there is such a distinction, but not, as far as I can see, any reason to believe that there is.

phenomenal quality – then we can avoid this aspect of the argument from queerness. However, we would apparently only do so by transferring the queerness from the supervenience relation (or the separate non-natural property of goodness) back to the hedonic tones themselves. There would, as far as I can see, be no reduction in the total amount of queerness.

What about epistemological queerness? It might be claimed that we have especially good epistemological access to our own experiences, and that our judgments about them are for that reason more trustworthy than judgments about other subjects. If it is true that pleasure is in some sense a direct experience of value, then this claim has some plausibility. There would, apparently, be no epistemological queerness found in judgments about hedonic value that is not also found in any other judgments about one's own phenomenal states. However, it would once more seem as if the queerness is not dispelled, but only included in the notion of hedonic tone. We appear to have made little overall progress.

Assume, on the other hand, that there is, at least in principle, a clear distinction between the phenomenal quality and the goodness or badness of hedonic states. Then the fact that we have a form of direct access to the phenomenal quality we are experiencing no longer means that we have the same special access to its goodness or badness. If it seems mysterious how we come to know that a given set of descriptive properties somehow ground or instantiate intrinsic (dis)value, I see no reason to think that mystery is to any extent dispelled if the descriptive properties in question are phenomenal properties with which we are intimately acquainted. Similarly, it is not as if we could reduce the epistemological queerness associated with the belief in the intrinsic badness of promise-breaking by claiming that we know the descriptive properties of promise-breaking really, really well.

As a last resort, one might argue that phenomenal qualities are very queer entities in the first place, and so it might not be too surprising that they should exhibit some further mysterious properties. This is what Mendola does: "Mackie's famous queerness objection seems to this degree correct: No ordinary machinations of the physical stuff of which we believe at least most of the world to consist can generate normative properties. Something as queer as the phenomenal must be considered before objective value becomes plausible." (Mendola 1990, 712) I think the force of this reply is limited. There is a difference between the queerness of what Nagel calls the subjective point of view (see e.g. Nagel 1986) and the queerness of independent authority. It is something it is like to have an experience of yellow, but there is no ultimate reason to seek this experience (considered apart from any hedonic reactions it might provoke). There is no obvious link between the two forms of queerness; it

seems, for instance, that we could easily imagine creatures with phenomenal consciousness who never felt good or bad. Thus, it is unclear how locating the normative queerness in an entity that also exhibits phenomenal queerness is supposed to solve any problem.

I think, then, that the existence of intrinsically reason-providing experiences would be quite perplexing. Whether this should get us to conclude that there are no such experiences depends on whether there is any alternative that is less perplexing. In the rest of this section, I examine some of the possibilities.

Authors sometimes think they can avoid ascribing realist (dis)value to hedonic states because they rely on what are – if my arguments in chapter 3 are sound – mistaken views on the nature of pleasure and pain. First, some confuse what I in section 3.1 called pain_s and pain_n : they mistake pain in the sense of suffering with physical pain or nociception. Since pain_n is not intrinsically bad, they think that the same goes for everything philosophers have discussed under the heading of “pain”. In *Reasons and Persons*, Parfit writes: “Some have claimed that pain is intrinsically bad, and that is why we dislike it. As I have suggested, I doubt this claim. After taking certain kinds of drug, people claim that the quality of their sensations has not altered, but they no longer dislike these sensations. We would regard such drugs as effective analgesics.” (Parfit 1984, 501; see also Goldman 2008, 520-1; Street 2006, 147-50) However, what is plausibly taken to be intrinsically bad is of course not the physical sensations these people are having, but their negative hedonic reactions to them. The drugs do not stop these reactions from being bad; they prevent them from occurring in the first place. There are, as far as I know, no drugs that remove the badness from suffering.

Second, one might be led to deny that there are hedonic ultimate reasons if one subscribes to the preferential or motivational view on pleasure and pain. Imagine that you are about to undergo long and complicated surgery, and that doctors are short on anesthetics and decide to just tie you up tightly and put duct tape over your mouth, so that you won’t disturb them. If the preferential or motivational view is correct, what happens in this case is that you have a sensory experience which is, in itself, neither good nor bad. You then form either an intense dislike of this sensation or a strong motive to escape it. Neither of these responses involves any intrinsically bad feel. I do not, as explained in chapter 3, believe that this is a satisfactory account of what would happen to ordinary human beings in a case like this. However, if this really is all there is to it, then I can understand why one would conclude that, as seen from a philosophical point of view, pleasures and pains do not really matter. For instance, the following remarks from Richard Kraut seem to make good sense, given this understanding of hedonic states:

even though we all want to avoid pain, I see no feature of it that makes it worthy of avoidance. We don't notice any characteristic of pain that grounds our aversion to it; we just hate the way it feels. But [...] just as our going for something does not show it to be good, so our avoiding it does not show it to be bad. And the fact that we *all* avoid it, and instinctively so, does not show it to be bad either. (Kraut 1994, 46)

Kraut believes that there are non-hedonic ultimate reasons, and so he must think that we should, to the extent that it psychologically feasible, simply stop caring about whether we feel good or bad. Rejecting the existence of hedonic ultimate reasons need not have this consequence, however. If one thinks that there aren't any non-hedonic ultimate reasons, either, then there is presumably nothing wrong with seeking pleasure and avoiding pain; it is just that failing to do so does not involve acting contrary to an ultimate reason. I take it that this is how we are to understand Alan H. Goldman's rejection of the objective value of attitudinal pleasure:

Taking pleasure in something in [the attitudinal] sense is wanting it to continue, liking the thought of it, and so on, and any state or activity is pleasurable in which one takes pleasure. Similarly, finding a state painful is mainly wanting it to stop. The problem here for claiming that pleasure in this sense has objective value is that taking pleasure in something is not only a subjective state, but is clearly relative to the subject's desires, likes and dislikes – in short, a matter of subjective values. What has objective value has it independently of subjects' attitudes, but taking pleasure in something or finding it painful is just a matter of subjective attitude (Goldman 2008, 521; see also Street 2006, 151-2)

Although they haven't explicitly said so, I suspect that many anti-realists have reasoned in a similar fashion. They have thought that pleasures and pains only consist of some sort of ultimately arbitrary desire to have or not have a certain sensation or experience, and that they therefore are not the kind of entities that could plausibly be said to ground ultimate reasons for action.

The preferential and motivational views need not be taken to undermine the notion of intrinsic hedonic (dis)value. One might (as Goldman goes on to note) think that the whole consisting of having an experience that one likes or wants to retain is good, independently of whether *it* in turn is liked or sought. Derek Parfit is, as we have seen, a representative of this view. There is something very strange about the idea that although the sensation of being cut open is not intrinsically bad, and although disliking it does not itself have any intrinsically bad feel, one just has to avoid being in a state where one dislikes the sensations of being cut open. Also, it was argued in section 2.3 that those who defend this view seem forced to accept that

the liking or successful seeking of any object constitutes a whole with intrinsic value. For these reasons, I find it less plausible than anti-realism. I will not pursue this question any further, however, since I believe it is raised by a misunderstanding of the nature of hedonic states.

The interesting question – which I will be discussing in the rest of this section – is whether we can reject hedonic ultimate reasons if the hedonic tone view is right and hedonic states do involve characteristic phenomenal feels. One possibility is to simply deny that these feels appear to have intrinsic (dis)value. If so, it is doubtful whether we can accept a strongly pluralist version of the hedonic tone view, since there would seem to be no way of explaining why we put completely different phenomenal feels under the single heading of pleasure. More importantly, I do not know of any philosopher who has made this claim. I take it that the would-be anti-realist about hedonic value would not deny that the pleasure phenomenology has an immediate appearance of goodness, but instead try to give some debunking explanation of the appearances. If he is convinced by the heterogeneity objection, he might then, presumably, solve the unity problem by saying that pleasures are those phenomenal feels that involve an illusion of intrinsic value.

We must not forget how difficult this form of debunking is. Remember that according to anti-realism about hedonic value, there are, from a philosophical point of view, no normative differences between the various phenomenal qualities human beings have the capacity of experiencing. They are all entirely neutral; they all have the same status as we ascribe to, say, the sensory quality of whiteness. What the anti-realist needs to explain is how creatures that only experience neutral phenomenal feels come to mistakenly believe that some of the feels are intrinsically good or bad. How does the appearance of phenomenal goodness arise if there is no such thing as real phenomenal goodness? Unless we are told a bit more about how this is supposed to happen, the illusion seems even more queer and perplexing than the real thing.

Anti-realists rarely address this issue specifically, and it is far from clear that their general remarks about projection are of any help. Often, they suggest that normative projection occurs via *emotions* (see Mackie 1977, 42-3 and in particular Joyce 2006, 4.4). It would not be very plausible, however, to claim that basic hedonic value judgments can be debunked in this way. First of all, we do not tend to have emotional reactions to our own hedonic states, at least not with the kind of strength and frequency necessary for explaining why we accord these states intrinsic normative importance. Pleasures and pains are, after all, our most basic and primitive subjective responses. Second, and more importantly, emotional debunking fails to constitute a separate alternative to hedonic debunking. Emotions are accompanied – and presumably

partly constituted (see Russell 2003) – by affect. If we remove the pleasure and pain from guilt, anxiety, pride or compassion, what is there left to project? Of course, emotions typically involve certain positive or negative judgments, but such judgments are, on the anti-realist outlook, supposed to be the result of projection, not the source. We do not, as far as I know, experience any positivity or negativity that is purely “emotional”, understood as something over and above the hedonic qualities and the cognitive contents of our emotional episodes. Thus, emotions do not possess any independent projective power that could potentially be used in a debunking account of hedonic intuitions.

It might seem more promising to claim that we value pleasure because we are wired to seek it and mistakenly take this to show that it is worth seeking. Mackie puts great emphasis on motivational projection in his account of how people come to make the error of believing in realist values (Mackie 1977, 1.10), and though his discussion is not about hedonic value in particular, he may have thought that it would cover this case as well. It is no doubt true that human beings often end up valuing things that they are motivated to seek. However, motivation is, at least when it is strong, nearly always hedonically laden in some way; we feel good when we get what we want and bad when we don’t, and it is unpleasant to resist strong urges and pleasant to give in. Thus, we cannot automatically assume that it is the motivation per se that is the driving force in this process.

If motivational projection is to explain why we value pleasure and disvalue pain, it would have to be the case that urges or other conative states have the power to generate normative projection independently of the hedonic qualities accompanying them. The mere having of an impulse to do something would make us think that there is an ultimate reason to do it. Moreover, this could not just happen occasionally or in unusual circumstances. It would have to be an extremely robust phenomenon, since it would underlie our attitudes towards hedonic states, which in turn underlie pretty much all our normative attitudes.

There is, as far as I know, no empirical evidence that motivation by itself has this kind of dramatic impact on normative beliefs. It seems very unlikely that it does. I do occasionally experience urges that appear to be hedonically neutral. For instance, I sometimes have an impulse to kick small stones lying on the sidewalk or to jump when I am standing close to the edge of a cliff, even if there seems to be nothing hedonically positive about doing these things or giving in to these temptations. However, I find it hard to believe that this phenomenon of pure volition is the source of all or most of my normative conviction, including my immediate evaluations of phenomenal qualities. It doesn’t matter how much pure urge you add to a neutral phenomenal quality: the situation is too transparent, and it should, at least with a

minimum of reflection, be obvious that the resulting experience is entirely neutral. Hedonic appearances of value, however, are not like that; they refuse to go away no matter how much one considers the possibility that they might stem from motivational projection.

To illustrate my point, imagine some creatures that are capable of experiencing the same phenomenal states as we do. When they perceive something white, for instance, it feels exactly like it does when you and I do so. The only phenomenal difference between us and them is that they never experience any of the phenomenal qualities that we call hedonic tones. Assume, further, that these creatures have impulses to seek the phenomenal quality of whiteness, and to avoid the phenomenal quality of blackness. We must be careful to make sure that we do not imagine these impulses to be intertwined with what we call positive hedonic tone. They would either have to be pure urges of the kind just described, or they would be laden with the phenomenal qualities of whiteness and blackness – which, as I have said, are experienced by these creatures in exactly the same way as by us. Now suppose that the creatures are equipped with the same capacity for introspection and normative reflection as we have. Some of them even write PhD theses about the nature and value of their phenomenal states. Are we to believe that it would be impossible, or at least extremely difficult, for these creatures to introspectively determine that their experiences have no value? All their “hedonic” experiences consist of are whiteness, blackness, with the addition of some neutral urge. How could they possibly fail to notice this?

If we are to be deluded about the value of our own experiences, there has to be some further complication besides motivation that makes us go wrong. One possibility is that some phenomenal feels have an illusion of intrinsic (dis)value is built into their very nature. This would, I think, be even more queer and mysterious than the realist alternative. If one thinks it is unclear how a mental quality could, by the very way it feels, be worth experiencing, it is not exactly obvious how a mental quality, by the very way it feels, come to seem to be worth experiencing for its own sake even if it is not.

Another possibility is that the illusion does not stem directly from the phenomenal quality, but arises in some way when we evaluate it. Instead of saying that we value positive hedonic tone because we are wired to seek it, the anti-realist could say that we do so because we are wired to *like* it – even if there is, as seen from a philosophical point of view, nothing intrinsically good about the way it feels. On this account, the way the hedonic tones feel is in itself irrelevant for the genesis of the illusion. The glow or aura typically provoked by toothache could have felt just like the glow or aura typically provoked by orgasms, and we

could still have been under exactly the same impression that there is a strong reason to avoid toothaches.

Even if I think that Nagel may be too hasty in saying that we just couldn't be mistaken about the value of our own phenomenal states, it is certainly not easy to see how this would come about. In a world without intrinsic value, how do you trick a creature into experiencing some mental qualities as intrinsically valuable? There is, of course, no empirical support for the existence of the illusion-producing mechanism; in fact, unless we are given some details about how it is supposed to work, the present debunking claim appears to be impossible to verify or falsify even in principle. It is hardly the kind of claim that dispels the mystery surrounding hedonic states.

At this point, the anti-realist might be tempted to appeal to an evolutionary explanation. If we had been capable of realizing that there is, in the final account, nothing intrinsically good or bad about any phenomenal states, we might be less motivated to engage in behaviors that are crucial to survival and reproduction. Perhaps being under an illusion that some mental states are inherently good or bad is a more efficient way of securing evolutionary desirable conduct than a mere urge that strikes out of the blue and does not involve any appearance of desire-independent value. In "A Darwinian Dilemma for Realist Theories of Value", Sharon Street utilizes this idea when arguing for anti-realism about egoistic hedonic value:

The realist tells us that it is an independent evaluative truth that pain sensations (however he or she defines them) are bad, and yet this is precisely what evolutionary theory would have predicted that we come to think. [...] In order to explain why we came to think that these sensations are bad, we need make no reference whatsoever to the *fact* that they are bad; we need only to point out how it tended to promote reproductive success to *take* them to be bad (due to their connection between bodily conditions that tended to diminish reproductive success). (Street 2006, 151)

Street is no doubt right that a belief in a near-universal and persistent hedonic illusion cannot be faulted on evolutionary grounds. However, what is hard to understand is not that, once the illusion somehow comes into existence, it will spread in the population. It is, rather, how the illusion works in the first place. How is the overwhelming appearance that there is something intrinsically bad with the way it typically feels to have intense toothache generated if there are no intrinsic hedonic values and all phenomenal qualities are neutral? To say that an evolutionary story makes this intelligible is like saying that one can explain how time travel works by pointing out that it is conducive to reproduction.

I find, then, that in the case of reasons based on one's own pleasure and pain, anti-realism is even more mysterious than realism. At the very least, I want to maintain that, unless anti-realists are able to provide some further details and arguments, their outlook is not clearly preferable to the one I have described in this chapter. We do not presently know that we are innately and persistently deluded about our hedonic states. It is not unlikely that pleasure really is good and pain really is bad. If so, pursuing pleasure and avoiding pain is our best shot at doing what we all-things-considered ought to do.

In this chapter, I hope to have made a convincing case that hedonistic egoistic values and reasons are special. It is possible to explain why most philosophers believe in and disagree about many other reasons and values in a way that doesn't undermine our normative intuitions concerning our own pleasure and pain. Also, while egoistic hedonic ultimate reasons do appear to be somewhat queer, there is no preferable anti-realist alternative in this case. We seem to be forced to accept the intrinsic goodness of pleasure and the intrinsic badness of pain in much the same way as we are forced to recognize the existence of phenomenal consciousness: The universe would be much simpler and understandable if these things weren't part of it, but they are there nonetheless.

Hedonic egoistic value does not usually get the special treatment it deserves. While the reasons to seek one's own pleasure and to avoid one's own pain are often taken to be the most promising candidates for ultimate reasons (e.g. Baier 1958, 7.1; S. Rachels 2003; Street 2006, sect. 9), they are not considered to be different in kind from non-hedonic and non-egoistic reasons. Thus, authors who are sympathetic to normative realism do not generally stop to consider whether the position can be defended more easily if we limit our value claims to those concerning one's own phenomenal states.¹¹⁶ Similarly, anti-realists – projectivists included – do not find it necessary to provide any arguments that specifically target hedonic egoistic value; they seem to just assume that their general arguments apply to ultimate reasons across the board.¹¹⁷ In fact, the passages from Street and Goldman from which I quoted above are the only anti-realist treatments of hedonic value I have been able to locate in the literature.

¹¹⁶ An exception is Torbjörn Tännsjö's *Moral Realism*, where it is argued that moral observation only takes place during pleasant and painful experiences (Tännsjö 1990, ch. 3).

¹¹⁷ To be sure, some anti-realists are primarily interested in morality. As explained in section 5.3, if one thinks that morality has to be both authoritative and impartial, one can hold an error theory about morality even if one believes that there are non-moral ultimate reasons. However, I am not aware of any moral anti-realist who has actually held or defended such a position.

Independently of one's final verdict on hedonistic egoism, it should be acknowledged that the debate on normative realism and anti-realism would benefit from a clearer distinction between hedonic egoistic reasons and other reasons.

9. Objections to hedonism and egoism

In the two preceding chapters, I presented my argument for hedonistic egoism. Since there are no well-developed arguments of precisely this kind for the position in the literature, there are no direct objections to the argument to be found there, either. However, there are of course a number of objections to the substantive claims for which I have argued. In this final chapter of the thesis, I examine some of these objections.

There are three ways in which my version of hedonistic egoism could be mistaken. First, it could give the wrong account of hedonic egoistic reasons. Second, it could be wrong in denying that there are non-hedonic ultimate reasons. Finally, it could be false that hedonic ultimate reasons are all egoistic. I will deal with these possibilities in turn.

9.1 Objections to the classical conception of hedonic egoistic reasons

According to Sidgwick, a hedonistic egoist is a person who aims, with temporal neutrality, to attain “the greatest possible surplus of pleasure over pain” for himself (Sidgwick 1907, 121, 381). Sidgwick considered this understanding of egoistic hedonic reasons self-evidently correct. However, it is certainly not. In fact, many philosophers have disagreed with it.

First, one may, as Mill did, question the assumption that mere quantity is the only normatively relevant attribute of pleasures. Mill thought that it would be better to be Socrates than a pig even if the pig experienced more pleasure (Mill 1871, ch. 1, par. 5). In order to assess this claim, we need to be clear about on what grounds the former option is supposed to be better. In section 8.2, I criticized the idea that the value of a pleasure depends on its object. Thus, if Mill’s claim is to be correct, it would have to be on purely phenomenal grounds: it simply *feels* better to be Socrates than to be the pig even if the pig experiences more pleasure.

On the monistic hedonic tone view, this claim clearly makes no sense. However, on a moderately or strongly pluralistic view, it does. A pluralist could, as explained in section 4.1, hold that some hedonic tones are better than others even when we keep their intensity constant. If pleasures feel irreducibly different, there need not be a simple one-to-one relationship between the amount of pleasure and its value. In fact, the pluralist might even agree with the stronger claim made by Mill that no amount of pig pleasure could outweigh any amount of Socrates pleasure. As long as the only basis for this claim is how these pleasures feel, making it does not in any way involve going beyond a hedonist outlook on

reasons. Nor is the maximizing element in classical hedonism discarded; it is not denied that hedonic *value* should be maximized, but only that the sum of pleasure should be.

Roger Crisp suggests that Mill's notion of qualitative distinctions between pleasures should be understood in this way, and apparently endorses the idea that some pleasures can be qualitatively better than others in the present sense (Crisp 2006, 4.4). I do not think there are any distinctions of this dramatic kind between experiences that are good because of how they feel. I do not know what it is like to be a pig, but I do believe that my sexual experiences, which I take it would, according to Mill, contain pleasure of a low quality, typically feel better than most of my other experiences, including listening to a late Beethoven sonata or reading *Pride and Prejudice* for the first time (the examples are from Crisp 2006, 111, 115). We must bear in mind that to say that a pleasure is better in this context means that there is more reason to seek it for how it feels, not that it is more aesthetically pleasing or a clearer sign of human perfection. Drug addicts staggering about talking to themselves and laughing do not make a pretty sight, and I can understand why one would have a sentiment of mild contempt towards people who only listen to simple popular music. But I doubt that there is much less reason to seek what these people are feeling for its own sake than to seek for its own sake what I am typically feeling when I enjoy myself. However, if I should turn out to be wrong about this particular matter, it will not have any consequences for my defense of hedonistic egoism in this thesis.

Some philosophers hold that there may be reasons not to maximize the intensity of one's pleasures even apart from considerations of the kind just mentioned. Jens Timmermann writes that:

Sometimes we decide that pleasure is bad, or not worth having, not because of any extrinsic factor like moral, aesthetic etc. constraints but rather because one is experiencing enough pleasure to the point that more would in itself be undesirable. Strong sensual stimulation can be like that: intense food, being tickled, sexual pleasure. In these cases, the pleasure on offer, beyond a certain point, simply exceeds the limit of what we can bear. (Timmermann 2005, 144)

Timmermann seems to think that if these preferences are accepted as reasonable, normative hedonism must be mistaken. This is not true; hedonism could just be reformulated so that pleasure is no longer a teleological value, or something to be maximized (see Hills 2008). We would have reason to seek a hedonic level that is optimal, rather than one that is as high as

possible. It hardly follows from Timmermann's claims that other things than pleasures must have intrinsic value. In any case, I have a hard time understanding the phenomenon Timmermann describes, since he never makes it clear what he takes hedonic states to be. Perhaps all he means is that sensations (or other mental states) that bring us pleasure can bring us less pleasure, and eventually even pain, when they get intense enough. No hedonist would deny that this is a real phenomenon, and all the views on pleasure and pain discussed in chapter 3 could account for it. However, it does not, of course, have any bearing on the validity of the claim that one ought to seek as much pleasure as possible. In order for this claim to be false, it would have to be the case that we sometimes correctly judge that an experience contains too much positive hedonic tone (or involves too strong a pro-attitude, or generates too strong a motive for its continuation). I have little to say about such judgments except for confessing that I have never felt any inclination towards making them. I hope this is because they would all be mistaken, and not just because I haven't felt sufficiently intense pleasure.

Michael Slote also questions whether there is always a reason to seek as much enjoyment as possible. He focuses, however, on more mundane pleasures. Imagine, for instance, that you are not particularly hungry, but yet know that you would derive some enjoyment from drinking a coke and eating a candy bar. Assume, also, that doing so would have no bad long-term effects. Would you then just *have to* have a snack? Slote thinks that you could reasonably "turn down a good thing, a sure enjoyment, because you are perfectly satisfied as you are. Most of us are often in situations of this sort, and many of us would often do the same thing. We are not boundless optimizers or maximizers, but are sometimes (more) modest in our desires and needs. But such modesty, such moderation, is arguably neither irrational nor unreasonable on our part." (Slote 1989, 10)

Why would one make a claim like this? Perhaps it seems a bit harsh to accuse someone of being unreasonable or irrational on account of failing to have a coke and a candy bar. However, we must keep in mind that this accusation would amount to nothing over and above a claim that the person has ignored a not particularly strong reason for action. Also, accepting that this reason is there need not involve a belief that one would, in a situation like this, be highly motivated to act on it. When our hedonic level is reasonably high, we sometimes feel no immediate desire for improvement. Furthermore, I suppose that many people have developed a non-hedonic liking of moderation in the way outlined in section 8.1; being greedy can, after all, have bad consequences and is typically discouraged by parents and society in general.

It turns out that Slote is not skeptical of maximization for any of these reasons. Instead, he has failed to distinguish between maximizing as a feature of a theory of reasons and maximizing as a decision procedure or the basis of a character trait. Thus, he rejects maximizing theories of reasons because he thinks that people acting on them would appear “needy and somewhat desperate, as well as cramped and unspontaneous” (ibid., 43). However, hedonistic egoism does not entail that one ought to be engaged in maximizing one’s hedonic level to the point where one can never relax or be spontaneous, but only to the point where the hedonic benefits of being thus engaged are larger than the hedonic costs. There seems to be nothing obviously unattractive or counterintuitive about this claim.

Yet another objection to Sidgwick’s form of hedonism can be found in Parfit’s *Reasons and Persons*. When choosing between two possible lives, you ought, according to Sidgwick’s version of hedonistic egoism, to simply add together the values of each episode of pleasure in the respective lives and opt for the life where the sum is greatest. Parfit thinks that such “summative” theories of what makes a life go best are implausible:

Suppose that I could either have fifty years of life of an extremely high quality, or an indefinite number of years that are barely worth living. In the first alternative, my fifty years would, on any theory, go extremely well. [...] In the second alternative my life would always be, though not by much, worth living. There would be nothing bad about this life, and it would each day contain a few small pleasures. [...] I do not believe that the second alternative would give me a better life. I therefore reject the Summative Theories. (Parfit 1984, 498-9)

Presumably, this means that he also rejects the claim that there would be most reason to choose the second alternative. Now we must be careful not to perform the thought experiment incorrectly by imagining it to be filled with boredom and frustration over the fact that one’s hedonic level pleasure is never higher than slightly above the zero level. We should also keep in mind that all the long periods with a hedonic level close to zero that can be found in the lives of most people are, according to hedonism, barely worth living (see Tännsjö 2002). Furthermore, we may have doubts concerning whether and in what sense one would be one and the same person if one’s life lasted for, say, a billion years. If, for instance, Parfit’s view on the proper object of egoistic concern is correct, we would have to use a certain discount rate for future pleasures, and so in order to make the thought experiment work, the pleasures

in the long life would actually have to eventually become quite intense. Then it might no longer seem intuitively obvious that one ought not to prefer it.¹¹⁸

On the other hand, it may be that we are all Cartesian egos, or that having a continued capacity for consciousness is sufficient to extend egoistic reasons into the future without any discount rate. If so, the classical conception of hedonic reasons would entail that the long life is better, and even if one adjusts for all the possible complications noted, this will presumably still seem counterintuitive. We need to ask, then, whether this intuition should be taken as a reflection of some normative fact, or whether there is some debunking explanation of it to be found.

Since this intuition does not concern any particular non-hedonic or non-egoistic value, but only how values should be distributed in a life, there is no obvious reason why a hedonistic egoist could not accept it at face value and adjust his theory accordingly. However, one would then face certain questions. First, it is a bit unclear how human beings would have epistemological access (albeit perhaps not a very reliable one) to facts such of the type “if a life is a million times longer than an ordinary human life, it needs to contain at least 235 times as much pleasure in order to be more choice-worthy”. Also, it is doubtful whether this intuition is consistent with other intuitions we have about these matters. Rather than going into details, I will just point to the fact that population ethicists have not been able to come up with any intuitively pleasing solutions to the interpersonal version of this problem (see Parfit 1984, part 4; Ryberg, Tännsjö & Arrhenius 2010).

I find it more likely that the intuition described by Parfit has its origins in the automatic and affectively laden information-processing system described in chapters 7 and 8. I suspect that, when one performs Parfit’s thought experiment, what happens is simply that one feels good about the first life and the bad about the prospect of leading a long life “barely worth living”. As usual, the goodness and badness become projected onto the options themselves. One has no clear image of what the second life would be like; it is simply represented as “very long and unattractive”. Thus, it makes little or no difference to the intuitive reaction whether the life is supposed to last 100 000 years or 100 000 million years, nor does it matter whether one tries to imagine that the pleasures it contains are one millionth or one billionth as intense as the pleasures in the first life.

¹¹⁸ This presupposes that one is making an egoistic evaluation of the life. From an impartial point of view, it would not matter whether or to what extent the future person would be you.

This insensitivity to numbers should come as no surprise if the intuition is taken to be nothing over and above an automatic affective judgment. It is well known that the automatic system is not very good with numbers, and that this has an impact on choices people make regarding risk, insurance and gambling (see e.g. Loewenstein, Weber, Hsee, & Welch 2001; Rottenstreich & Hsee 2001, 185-190). Findings by Paul Slovic and his colleagues provide an elegant example of how affectively based judgments on quantitative issues can go astray. People who were offered a bet where they had a 7/36 chance of winning \$9 and a 29/36 of losing 5c tended to give a much higher rating of the attractiveness of the bet than people who were offered a 7/36 chance of winning and no chance of losing at all. Similarly, people who were asked whether they will support a safety measure that will save 98% of 150 lives at risk were significantly more positive than people who were asked whether they will support one that will save 150 lives. Slovic thinks the explanation is that while the benefit of winning \$9 or saving 150 lives has unclear affective value, we feel much better about the outcome when we compare the sum of \$9 with a much smaller one, or when we see a very high success rate such as 98% (Slovic 2002).

I believe, then, that Parfit is wrong to reject classical hedonism and other summative theories of well-being. It should be kept in mind, though, that this claim is not essential to my defense of hedonistic egoism.

I will discuss one more worry about Sidgwick's understanding of hedonic egoistic reasons. Let us say that the average human life contains 100 units of hedonic value. Imagine a life consisting only of one very long but unchanging episode of pleasure. In order to avoid complications about the quality of different pleasures, we can assume that the enjoyment in question is enjoyment of late Beethoven sonata (or whatever enjoyment it is that would be of the highest kind). If we make the pleasure intense enough, this life could easily be made to contain more than 100 units of hedonic value. In fact, since there is much pain and not all that much intense pleasure in human lives, the intensity would not have to be particularly high for the life to surpass 100 units of hedonic value. Would such a life really be better than an ordinary human life?

As with the previous objection, it is easy to imagine this life filled with boredom and frustration. In order for the thought experiment to work, however, we must assume that the person living it is not plagued by the lack of variety in his experiences, nor does he get tired of hearing the sonata. Every single time it feels reasonably good. Also, we must keep in mind that we are discussing the nature of hedonic egoistic reasons, and not the question of whether

these are the only reasons for action there are. If we are to conclude that the life imagined is not better than an average life, we must not do so because we think it is lacking in non-hedonic respects, but because it doesn't feel good enough for the person living it. Finally, it must be recalled that hedonism, as explained in section 4.2, takes a somewhat deflationary view of human existence: even an ordinary life is not all that great after all.

If the monistic hedonic tone view is correct, a supporter of hedonistic egoism seems forced to accept that a life consisting only of repeated moderate enjoyment of a late Beethoven sonata would be better than most actual lives. If there is an intuition to the contrary, it could presumably be debunked in the way that should by now be familiar: we enjoy novelty and variation, and so come to think of it as valuable in itself.

Given some form of pluralism about hedonic tones, it would, on the other hand, be possible to modify hedonistic egoism so that it accords a certain intrinsic importance to experiential variation, or holds that the value of a given pleasure with a given intensity increases in value more slowly the longer it has been felt. In order to defend such modifications, one would of course have to explain how the intuition underlying them is relevantly different from intuitions that do not, on a hedonistic egoistic outlook, deserve to be given epistemological credit. It is hard to see how such an explanation would go, but I will not pursue the matter any further here.¹¹⁹

9.2 Objections to hedonism

Hedonism about reasons is a very bold and sweeping view. If it could be shown that there is just a single ultimate reason that is not hedonic in nature, the view would have to be rejected. The literature abounds with examples and thought experiments that are meant to perform this feat by revealing clear and persistent non-hedonistic intuitions. Since all these intuitive objections raise roughly the same set of issues, I will only discuss one case in detail, namely

¹¹⁹ It has also been claimed that the shape of the well-being curve of a life influences how good the life as a whole is: we have reason to want a life that keeps getting better and better than the opposite, even if the two lives contain equally many episodes of equally intense episodes of pleasure and pain (see e.g. Velleman 1991). I do not, however, think that classical hedonism is particularly counter-intuitive on this point. As Feldman points out, a sense that one's life is improving is a source of pleasure, so we would have to make sure that the people living the lives in question do not notice the how their lives develop, or, more realistically, compensate for this. Moreover, we must remember that we are discussing non-derivative reasons for action and not what is an admirable or narratively pleasing life (Feldman 2004, 6.3).

the most famous and influential of them all. I am, of course, talking about Nozick's experience machine:

Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life's experiences? [...] Others can also plug in, so there's no need to stay unplugged to serve them. Would you plug in? (Nozick 1974, 42-5)

According to hedonism, doing so is apparently a very good idea.¹²⁰ However, the immediate intuitive reaction is of course that one would not. Many philosophers have taken this as conclusive (or at least very strong) evidence that hedonism is mistaken (see e.g. Griffin 1986, 9-10, 19-20; Sumner 1996, 94-8). In this section, I consider various replies that could be made on behalf of hedonism.

Those who question Nozick's objection typically start out by pointing to ways in which the thought experiment can go wrong. For instance, the reluctance to plug in may stem from an aversion to put oneself completely at the mercy of other people, or a fear that the machine may malfunction (Shaw 1999, 50-1, Hewitt forthcoming, sect. 2). Both these reasons for rejecting the experience machine are, of course, compatible with hedonism. In order to avoid these complications, we can follow a suggestion from Roger Crisp and instead focus on two parallel lives which are, throughout, introspectively indiscernible. The first is lived in the real world and contains much genuine accomplishment, the other is lived in an experience machine (Crisp 2006, 118). Isn't it clear that there would be a reason to prefer the former life?

A hedonist may of course simply deny that he has any intuition to this effect. This is the line taken by Torbjörn Tännsjö (Tännsjö 2007, 92-5; he also suggests that many ordinary people would share his view). I have some sympathy with this reply; after thinking about these issues for some time, there is not much left of what used to be a quite powerful non-hedonistic intuition about the value of authenticity. However, it would seem that the

¹²⁰ If hedonism is understood as a theory of well-being rather than as a general theory of reasons, a hedonist might claim that there would be weighty reasons that do not concern well-being, such as aesthetic or perfectionist considerations, not to plug in. Nozick's objection may then seem to have considerably less force (see Kawall 1999, sect. 3; Silverstein 2000). However, this kind of hedonism would still be committed to the claim that whether one is connected to an experience machine makes no difference to one's well-being, and most philosophers would find this claim, too, counterintuitive (see e.g. Griffin 1986, 19-20).

explanation of this fact need not be that I have become a better intuitive knower of normative facts. It could be that my general commitment to hedonism has clouded my judgments. More relevantly, it appears that a clear majority of professional philosophers do have the intuition that some form of authenticity or accomplishment matters for its own sake. Unless some positive reason is given for thinking that there is something wrong with this intuition, it is hard to see why it should be reasonable to side with the minority. At the very least one should suspend judgment.

Most critics do not deny that Nozick's example reveals a reasonably strong intuition about the intrinsic value of authenticity or accomplishment. Instead, they try to debunk this intuition in various ways. According to D. W. Haslett, the problem is that Nozick's case is too "fantastic" for our intuitions regarding it to be trusted:

in reality, being a floating brain [...] would no doubt make a difference in our experience that was both extensive and undesirable. We may know perfectly well that we are to hypothesize that, contrary to reality, these things would not make a difference. Yet our emotional reactions to these things will naturally tend to reflect reality rather than this extraordinary hypothesis. And, naturally, our *intuitions* about whether these things make us worse off will tend to reflect our emotional reactions. That is why, with these kinds of cases, our 'intuitions' tend not to be reliable guides. (Haslett 1994, 40)¹²¹

I find it hard to believe that this could be the correct explanation of why Nozick's objection has been so influential. Though it may be unclear precisely how the experience machine is supposed to work, it does not seem particularly difficult to envision what it produces, since it would be just like something with which we are very closely acquainted, namely our own experiences. Insisting that those philosophers who agree with Nozick are not able to imagine their experiences being caused by an experience machine without tacitly assuming that they would somehow feel different seems both implausible and condescending.

Joseph Mendola questions the negative intuitive evaluation of a life in the experience machine in a more indirect way. He observes that there may be different views on how the importance of authenticity depends on one's preferences and how it compares to the importance of other values, and that none of these views are unequivocally favored by common sense. This should make us wonder whether our non-hedonistic intuitions are reliable trackers of genuine normative facts (Mendola 2006, sect. 2). Mendola may be right

¹²¹ See Hewitt forthcoming, sect. 2 for a similar but more elaborate objection.

about this, but if the existence of intuitive disagreement regarding the precise nature of the value of authenticity should get us to reject the whole idea that authenticity is intrinsically valuable, then the existence of intuitive disagreement on the precise nature of hedonic reasons should get us to reject the whole idea that pleasure has intrinsic value. In fact, Alan H. Goldman objects to the idea that pleasures ground ultimate reasons precisely on the grounds that it is intuitively unclear how different pleasures should be evaluated and compared (Goldman 2008, sect. 5). Surely Mendola and Goldman both jump to conclusions. If we think that normative theory should be guided by a general reliance on intuition, the reasonable thing to say would be that while it is obvious that pleasure and authenticity give rise to ultimate reasons, it is not obvious what the exact contents of these reasons are.

In his discussion of the experience machine objection, Roger Crisp points out that a hedonist can in a sense recognize accomplishment as something that should be pursued for its own sake, since doing so would presumably be hedonically desirable: “If we allow that in the usual case someone will enjoy accomplishing more than accomplishing less, then there are good reasons to think that motivation by non-hedonist beliefs may be more successful, by hedonist lights, than motivation by hedonist beliefs.” (Crisp 120; see also Hewitt forthcoming, sect. 4) There is undoubtedly some truth to this observation (though as argued in chapter 4, it is not clear that it is necessary to have non-hedonist beliefs in order to be non-hedonistically motivated). However, it seems to have little bearing on whether what hedonism says about accomplishment is correct. The intuitive judgment provoked by the experience machine case is not a judgment that we should act as if accomplishment has intrinsic value, or that it would be a good thing if we falsely believed that accomplishment has intrinsic value, but a judgment that it really has intrinsic value.

Crisp offers another suggestion that appears more promising. “Could it not be”, he asks, “that our valuing of accomplishment is an example of a kind of collective bad faith, with its roots in the spontaneous and largely unreflective social practices of our distant ancestors?” (op. cit., 121-2) However, in Crisp’s discussion, this remains just a suggestion.¹²² We are not given any account of the mechanism that is supposed to have generated this bad faith, nor any reason for why this debunking account should be preferable to a vindicating evolutionary account where our normative beliefs are a (by-)product of our evolutionarily useful capacity

¹²² The reasons for being sceptical of non-hedonic intuitions given by Sharon Hewitt in her paper “What do our intuitions about the experience machine really tell us about hedonism?” are similarly vague and unspecific (Hewitt forthcoming, sect. 3).

for rational insight. Finally, we are not told why it is that the valuing of enjoyment would not be susceptible to a similar form of debunking. Without such elaborations, the mere fact that a debunking explanation of non-hedonistic value judgments appears possible is not a good reason for holding that hedonism is a plausible normative theory.

Matthew Silverstein makes a more sustained attempt at explaining away the non-hedonistic intuition revealed by the experience machine case. He argues that while our desires do not typically have pleasure or happiness as their object, we nevertheless tend to end up desiring things that have brought us pleasure. Thus,

the desire to track reality owes its hold upon us to the role it has played in the creation of happiness. We acquire our powerful attachment to reality after finding again and again that deception almost always ends in suffering. [...] Our intuitive views about what is prudentially good, the views upon which the experience machine argument relies, owe their existence to happiness. We miss the mark, then, if we take our intuitions about the experience machine as evidence against hedonism. (Silverstein 2000, 293-6)

Silverstein's account of how desires are generated is similar to that given by Mill and appears a bit simplistic, but let us disregard that here. The real problem is that Nozick's example does not appeal to what we desire, but to what we, on intuitive grounds, believe to be *desirable*. In his book *The Examined Life*, Nozick makes this clear:

Notice that I am not saying simply that since we desire connection to actuality the experience machine is defective because it does not give us whatever we desire [...] Rather, I am saying that the connection to actuality is important whether or not we desire it – that is why we desire it – and the experience machine is inadequate because it doesn't give us *that*. (Nozick 1990, 106-107)

Of course, if it could be shown that our intuitive non-hedonic beliefs are also dependent on hedonic experiences, Silverstein would have a very good argument. However, he offers no justification for thinking that they are other than a brief remark that “our intuitions tend to reflect our desires and preferences.” (op.cit., 293) Maybe they do, but many authors have thought that the connection often goes in the opposite direction: we tend to desire what we intuitively (or reflectively) take to be worth desiring. It is very unlikely that Nozick and the philosophers who have followed him would, without further argument, be prepared to consider their non-hedonic intuition a mere reflection of what they happen to like or want.

I find, then, that the reasons that have been given in the literature for resisting Nozick's argument are generally weak and unconvincing. If hedonists have no better replies to offer, we should take the experience machine case and similar objections to count significantly against their theory. I believe, of course, that there is a better reply available. As the reader might have guessed, it relies on the account of hedonic projection presented in the previous chapter.

We really do have non-hedonic intuitions about accomplishment and authenticity, and these intuitions present themselves as genuine insights into what there are desire-independent authoritative reasons to seek, and not as mere wants or likes. Still, it may very well be the case that hedonic experiences play a crucial part in the explanation of why we have these intuitions. Our automatic cognitive system has registered that accomplishment and authenticity generally are accompanied by high hedonic levels and, perhaps even more importantly, that failure and deceit makes us feel bad. In particular, the idea that one may not be as talented as one thought, or that other people may not really like one as much as they pretend to do, poses a very unpleasant threat to one's self-esteem. These tendencies are then reinforced through evaluative conditioning: parents tend to tell their children that only authentic accomplishment is praiseworthy, and presumably we also to some extent associate these values with other things we value for their own sake.

All this happens automatically and long before we start reflecting philosophically on our conduct. It is not as if we all go through a period of pursuing accomplishment and authenticity explicitly as means to pleasure, and then gradually develop an independent desire for these things, in the way one might develop an independent desire for going on vacation if previous vacations have been enjoyable. Thus, we have no introspective access to the dubious causal history of the intuition, and that is why most philosophers have not found any particular reason to question it.

This debunking explanation of the experience machine intuition may seem very speculative. There is, of course, no empirical data on this particular issue. We should not forget, however, that it is an established fact that evaluative conditioning takes place, and more generally that our practical reasoning is influenced by affective reactions in important but often unnoticeable ways. It would be very surprising if processes of the kind just sketched had absolutely no effect on our intuitive evaluations of authenticity and accomplishment. Given all the problems we would face if we take the experience machine intuition to reveal an objective normative fact, I actually find rejecting the epistemological credibility of this intuition to be the more straightforward and down-to-earth than taking it at face value.

The lesson that can be learned from this discussion of the experience machine also applies, I believe, to other intuitive objections to hedonism. When hedonists defend themselves against counterexamples to their theory, they tend, first, to overestimate how difficult it is to conduct thought experiments. Second, they have too little confidence in the analytic skills of their non-hedonist colleagues. Finally, they are too cavalier in their dismissal of non-hedonic intuitions as mere emotional or conative phenomena, perhaps with some dubious evolutionary history. If hedonism is to be a credible view, we need a clear and empirically supported account of how non-hedonistic intuitions go wrong and a principled explanation of why these intuitions are different from the intuitions that pleasure is good and pain is bad. This does not, of course, have to be done precisely in the way I have done it, but it is hard to see how a debunking of non-hedonic intuitions could take a completely different form. In other words, it seems that if we are to hold that hedonism is correct, it would have to be because of considerations of the kind offered in this thesis.

9.3 Nagel's objections to egoism

In section 5.2, I reviewed some arguments that were meant to establish that egoism is incoherent or that anyone who believes in authoritative reasons for action is implicitly committed to recognizing certain non-egoistic reasons with independent authority. However, most of the philosophers who believe that egoism is mistaken do not do claim to be able to prove this in any strict sense. Still, they have offered various considerations that are meant to throw the plausibility of the theory into doubt. I have already explained (in section 8.2) why I think that an egoist need not worry about simple appeals to non-egoistic intuitions (such as those made in Shafer-Landau 2003, 196-9; Crisp 2006, 131-2). In this section, I want to discuss two different objections to egoism, both from Thomas Nagel. First, I assess his claim that egoism is incompatible with some of our emotions and reactive attitudes to the point that it is almost psychologically impossible to be an egoist. Second, I criticize the influential argument in *The View from Nowhere* which aims to show that egoism is in conflict with the way hedonic states are subjectively experienced.

I will, unsurprisingly, argue that Nagel's objections to egoism are not convincing. It should be pointed out that I do not just claim that an egoist need not be convinced by them if he is sufficiently stubborn and single-minded, or that there are or could have been people of a very unusual kind over which the claims of other people had no independent normative authority. My view is that no-one should find the anti-egoistic arguments convincing, and that

any human being who is taking an interest in morality or other people beyond what rational egoism recommends is acting contrary to his ultimate reasons.

Nagel's hostile attitude towards egoism seems in part to derive from a certain picture of what it would be like, psychologically, to be an egoist. This picture is so peculiar that Nagel thinks that even people who profess to be egoists will fail to fit it (Nagel 1970, 84). More precisely, the alleged problem is that an egoist is forced to have a split attitude towards his own suffering: on the one hand, he thinks it is an objective fact that he himself has a strong reason to avoid it; on the other hand he must admit that it is, by itself, normatively irrelevant for other people. Thus, it will seem unintelligible if anyone else should have a non-derivative desire for it to go away:

Suppose I have been rescued from a fire and find myself in a hospital burn ward. I want something for the pain, and so does the person in the next bed. He professes to hope that we will both be given morphine, but I fail to understand this. I understand why he has a reason to want morphine for himself, but what reason does he have to want *me* to get some? Does my groaning bother him? (Nagel 1986, 160)

Similarly, the egoist would seem to have no basis for expecting or demanding that others should refrain from making him suffer. He is, in other words, "precluded from feeling resentment, which embodies the judgment that another is failing to act on reasons with which one's own needs provide him. No matter how extreme his own concern the egoist will not feel that this in itself need be of interest to anyone else." (Nagel 1970, 85, see also 82-3; Darwall 2006, 74-9)

In the passages from which I have quoted, Nagel seems to assume that when one believes in a theory of reasons, all one's emotions and reactive attitudes will be modified accordingly. The fact that an egoist does not believe that his needs are a non-derivative source of reasons for action to other people is automatically taken to show that he cannot *feel* as if they are. There is, as I have explained earlier in this thesis, no such direct relationship between our abstract philosophical beliefs and our psychological make-up. Consider envy, for instance. By Nagel's reasoning, one could not feel envy without having made a judgment that there is an ultimate reason for preventing the person whom one envies from succeeding. Surely we do not typically believe there is such a reason; we just find ourselves wanting to do so anyway.

In section 6.2, I argued that believing in egoism does not by itself make one incapable of feeling sympathy, and that most people would find it in their best interest to keep and perhaps

even strengthen their sympathetic bonds to other people. Thus, it would not be particularly surprising to the egoist if his roommate wants him to get some morphine. Moreover, the roommate's hope is not to be understood as a desire for the rush of sympathetic pleasure that he would feel if this were to happen, but as a non-derivative hope that is automatically formed simply because of how the human mind works. Similar remarks apply to resentment. No philosophical belief has the capacity to preclude us from finding hardship that is brought on deliberately by other people especially difficult to bear, or to turn the opposition we feel towards morally bad people into the same kind of opposition we feel towards earthquakes or malfunctioning computers.

To be sure, I do not want to claim that a believing in and acting on egoism has absolutely no effect on one's reactive attitudes. My point is just that the effect would be rather limited, and that there is nothing obviously wrongheaded or unattractive about how a typical egoist would *feel* about other people. One could, of course, define the reactive attitudes in such a way that they must include a philosophical belief that they are justified by authoritative reasons for action. If so, an egoist could not have these attitudes. However, he could have some very similar attitudes, which may even include the appearance of authoritative reasons for action. The only difference would be that the egoist believes that, ultimately, this appearance is illusory. It is hard to see how this psychological make-up would be so peculiar that one would "doubt whether there are any genuine specimens of the type" (Nagel 1986, 84).

One might argue that even if an egoist could have much the same feelings concerning other people as a non-egoist, the nature of these feelings nevertheless suggests that the egoist is mistaken. We should not, however, assume that our feelings must correspond to or be justified by ultimate reasons. Surely there are a number of psychological and evolutionary considerations that could explain why we have special emotional responses to the conduct and well-being of other people. If a would-be egoist need not worry about non-egoistic intuitive judgments, he need not worry about non-egoistic emotions, either.

In section 8.2, we saw that if one wants to argue that non-egoistic hedonic intuitions are resistant to hedonic debunking, it is not enough to point out that they are stronger or more prevalent than the various non-hedonic intuitions that can be explained away in this manner. One would need to show that they have a different causal history, or stem from a different cognitive mechanism. In a short, but often approvingly cited passage (see e.g. Parfit forthcoming, 149), Nagel tries to do this. In fact, he argues that intuitive non-egoistic

convictions concerning pain have precisely the same source as the conviction that pain is bad in the first place, namely how it unreflectively feels to experience it. If this is true, then my defense of egoism in this thesis obviously breaks down. I do not think it is true, and I will explain why in the remainder of this section. However, it should be kept in mind that if I am wrong, then this would not affect my defense of hedonism.

Nagel starts by observing that a pain “can be detached from the fact that it is mine without losing any of its dreadfulness.” (Nagel 1986, 160) If this is taken to mean that pain has the same awful phenomenal quality regardless of whose pain it is, then this observation is presumably correct, but not relevant to the present issue, since it is not in conflict with egoism. Surely an egoist knows that other people feel pain the same way as he does; it is just that he believes that pain is only a source of ultimate reasons to the person who is experiencing it.

Nagel goes on, however, to make the same claim about the *badness* of pain. He thinks that the “objective self”, which considers the world from no particular point of view, will still find that pain experienced by any given subject is a bad thing. There is of course a sense in which even a hedonistic egoist can say that, as seen from an impartial view, it is bad that people suffer, or that reducing the amount of suffering makes the world a better place. This kind of value would, however, not be a non-derivative source of reasons for action for any individual agent. It would be a mere sum of the things that provide reasons to individual agents possessing them.

What Nagel wants to say, however, is that the pain of any given individual grounds an agent-neutral reason, i.e. a reason with authority over all agents. It would seem that whether one thinks that the objective self will see that there are such reasons depends simply on whether one has a prior conviction that such reasons exist. It is at this point that Nagel appeals to the phenomenology of pain experiences:

When the objective self contemplates pain, it has to do so through the perspective of the sufferer, and the sufferer’s reaction is very clear. Of course he wants to be rid of *this pain* unreflectively—not because he thinks it would be good to reduce the amount of pain in the world. But at the same time his awareness of how bad it is doesn’t essentially involve the thought of it as his. The desire to be rid of pain has only the pain as its object. This is shown by the fact that it doesn’t even require the idea of oneself in order to make sense: if I lacked or lost the conception of myself as distinct from other possible or actual persons, I could still apprehend the badness of pain immediately. [...] My objective attitude toward pain is rightly taken over from the immediate attitude of the subject, and naturally takes the form of an evaluation of the pain itself, rather than merely a judgment of what would be reasonable for its victim to want: “*This experience* ought not to go on, *whoever* is having it.” To regard

pain as impersonally bad from the objective standpoint does not involve the illegitimate suppression of an essential reference to the identity of its victim. In its most primitive form, the fact that it is mine—the concept of myself—doesn't come into my perception of the badness of my pain. (ibid., 161)

As I have said, what Nagel wants to establish is the existence of agent-neutral hedonic reasons. His introspective argument is mostly formulated in terms of “badness”, and so he must be assuming that what is bad must also be a source of reasons. As noted in section 8.2, it is certainly possible to question this assumption, at least when it comes to hedonic value. Let us, however, grant that these values do have the ability to ground reasons for action.¹²³ It is still not clear to whom something that is hedonically bad is supposed to provide reasons. This ambiguity makes Nagel's argument seem more plausible than it is. If we make the reasoning more explicit, we will see that there is no version of the argument which has plausible premises and at the same time leads to a non-egoistic conclusion.

One possibility is that the badness that is immediately apprehended is supposed to constitute a reason for all agents. This would give us the following argument:

- (1) When I experience pain, I immediately apprehend it to be bad in the sense that there is a reason for any agent to end this particular pain.
- (2) This apprehension does not require that I have the concept of myself or think of the pain as mine.
- (3) Therefore, egoism is probably false.

This is clearly not a very good argument. (1) is doing all the work here and (2) is not needed to reach the conclusion. If this is his argument, Nagel would simply be insisting from the outset that we directly apprehend that there are agent-neutral hedonic reasons; whether I happen think of the pain as mine or not would be beside the point. Even a non-egoist should admit that (1) is implausible as it stands.

Here is another possible argument:

- (1a) When I experience pain, I immediately apprehend it to be bad in the sense that there would be a reason for any person having this experience to end it.

¹²³ If they don't, one could give an argument of Nagel's kind for the claim that the present hedonic levels of all people at least matters to a given agent right now. I will not discuss this argument, but I believe it could be criticized in the same way as the argument involving reasons.

(2a) This apprehension does not require that I have the concept of myself or think of the pain as mine.

(3a) Therefore, egoism is probably false.

If this is how we understand Nagel's argument, the conclusion obviously doesn't follow from the premises. Egoism is entirely compatible with the possibility that one can focus on the phenomenal quality of pain, entirely disregarding the fact that it belongs to oneself, and then immediately see that anyone who had this experience would have a reason to end it.

The reason judgment in (1a) may seem a bit too sophisticated for it to be the object of an immediate apprehension. As noted in section 8.2, it is possible that we at first just become aware of the fact that the pain has some sort of normative status, and that we do not directly know for whom it has this status. This would give us the following argument:

(1b) When I experience pain, I immediately apprehend it to be bad in the sense that it gives rise to an authoritative reason to end it.

(2b) This apprehension does not require that I have the concept of myself or think of the pain as mine.

(3b) Therefore, egoism is probably false.

I have no fixed view on whether the badness of pain is initially apprehended in this impersonal fashion. If it is, that would of course not rule out egoism, since the truth of egoism is compatible with what is apprehended in (1b).

If Nagel's argument is going to succeed, I would have to be capable of apprehending that I have a reason to alleviate my pain, without thinking of the pain as mine. Thus, the argument would have to look like this:

(1c) When I experience pain, I immediately apprehend it to be bad in the sense that there is a reason for me to end it.

(2c) This apprehension does not require that I have the concept of myself or think of the pain as mine.

(3c) Therefore, egoism is probably false.¹²⁴

¹²⁴ In "Nagelian Arguments against Egoism", Stuart Rachels agrees that that this is the rendition of Nagel's argument that would yield a non-egoistic conclusion (S. Rachels 2002).

However, given that (1c) is accepted, (2c) begins to look quite dubious. It is clear that I would at least need to have a concept of myself in order to think that I (as supposed to merely someone) have a reason to end the pain. The question is whether I can apprehend this reason without at the time thinking of the pain as mine. Can the ownership of the authoritative hedonic reason be established before or independently of the ownership of the experience? It is important to stress that this must not be understood as a thought experiment. Nagel's claim is not that if I abstractly imagine one of my pains just floating around from person to person, so to speak, I would see that I have a reason to end it. That would just be an appeal to an intuition of the kind that could stem from hedonic projection. Rather, Nagel would have to be saying that there are actual cases where an agent does not even implicitly rely on the fact that the pain belongs to him and still immediately knows that there is an authoritative reason *for him* to end it.

At this point, the time for arguments is over, and the matter must be decided by introspection. I do not think that I have ever apprehended that there is a reason belonging to me for ending one of my pains without this having anything to do with the fact that the pain is mine. This is not because I find it impossible to focus exclusively on the phenomenal character of pain and bracket the fact that it happens to be experienced by me. This I may be able to do. However, to the extent that I succeed in this bracketing, the resulting reason judgment also seems to lose its reference to myself, and the reason becomes just a reason for "the person experiencing this", or perhaps just for "someone". In other words, while I do think I have introspective evidence that the phenomenal quality of pain generates authoritative reasons whenever it is experienced, I do not find any evidence that any particular instantiation of it gives me a reason independently of the fact that it is mine. I cannot, of course, pronounce on what others experience.¹²⁵ The reader will have to judge for himself.

Since introspection is difficult and may not yield conclusive results, it is perhaps worth the while to say a few words about the normative consequences of accepting Nagel's argument. If I, when experiencing pain, directly apprehend that pain has normative authority over me completely independently of whose pain it is, then it is hard to see how we could avoid a strict form of impartialism about hedonic reasons. We may have some pre-theoretical convictions to the effect that there is a special ultimate reason to be concerned with one's own hedonic states, but if these anti-impartial convictions are in conflict with the very way pain is

¹²⁵ Stuart Rachels claims to be able to see that he has a reason to end a pain without relying on the fact that it is his (S. Rachels 2002, sect. 4).

experienced, then it would seem that they would have to be discarded. What sort of egoistic or non-impartial hedonic intuitions could possibly override the way it subjectively feels to be in pain?¹²⁶

In fact, Nagel does embrace strict impartialism about hedonic reasons. At first sight, it might seem as if he does not, since he conceives of morality as a compromise between agent-relative and agent-neutral reasons (Nagel 1986, 195-204). However, the need for such a compromise does not, on Nagel's view, stem from the fact that agent-relative hedonic reasons have independent normative authority. Rather, the problem is that typical human beings are motivationally incapable of doing what they have most hedonic reason to do. We must thus "strike a balance between our higher and lower selves in arriving at an acceptable morality." (ibid., 202) However, it remains true that an ideally rational person would do more than morality requires. Those who perform supererogatory actions "are praiseworthy for submitting themselves to the *true* strength of reasons that they could not reasonably be required to follow strictly, given the mixed character of human motives." (ibid., 204, emphasis added)

I find Nagel's outlook on reasons implausible for two reasons. First, it is a very extreme normative view. Let's say that an evil demon approaches you and tells you that he has decided to make one person suffer excruciating pain after his death. The person is either going to be you or someone whom you don't know and will never meet. Assume also that the demon will make you forget that you have ever talked to him. The demon would torture you for 50 years and the other person for 51 years. According to impartialism about hedonic reasons, volunteering for the torture is what you ought to do – and not only morally, but all-things-considered. In fact, not doing so would be a serious mistake – more serious than any normative mistake an ordinary human being is likely to make in his life – since it leads to a whole year of unnecessary intense suffering. Even if Nagel thinks that human beings can be excused for such mistakes because of their "motivational complexity", it remains a mistake nonetheless. If this is what follows from Nagel's claims about how the badness of pain is subjectively apprehended, one might perhaps want to re-examine these claims.

¹²⁶ Conceivably, it could be argued that the way it subjectively feels to be in pain itself suggests some sort of compromise between egoism and strict impartialism. This claim does not seem to have much initial plausibility, and I am not aware of any attempts at defending it in the literature. (There are of course authors who defend middle positions between egoism and strict impartialism (see e.g. Crisp 2006, 5.2-5.3; Parfit forthcoming, ch. 6), but their views seem to be based not directly on the phenomenology of pain, but on the kind of intuitions criticized in section 8.2.)

Second, I think that Nagel overestimates the extent to which human beings are incapable of doing what they think they have most reason to do. He is not only the only philosopher to do so.¹²⁷ Although I haven't made any careful empirical studies on this matter, I suspect that many normative philosophers hold views about non-egoistic reasons that they do not act on in everyday life, and that they would cite their motivational shortcomings as the explanation for this discrepancy. I can see why one would find an outlook like this attractive: one gets to hold a very nice and noble theory of reasons without having to be all that nice and noble oneself. However, it does not quite add up.

I do not want to deny that many impartialist reason judgments would be hard to act on even if one thought they were valid. Even something as little intimidating as visiting an old aunt from time to time might take quite a lot of willpower if one is busy. However, not all normative choices concerning oneself and other people are like this. Assume that you are a hedonic impartialist of some kind and you think that you ought to give, say, 10% of your salary to charity. This is not a difficult thing to do. It is not as if you would sit down in front of the computer and try to make the transaction and then just become overwhelmed with self-love or the desire for a new TV or whatever you would want to spend the money on. Of course I am not denying that even philosophers may sometimes suddenly find themselves wanting something so much that they are strongly tempted to buy it even if they think there is most reason not to. I am just claiming that it is not like this all the time. If heroin addicts are able get rid of their addiction by signing up for therapy, then surely intelligent and reflective people like philosophers are able to give as much money to charity as they think they ought to give (or at least more than they typically do).

Even in an extreme situation like the demon case described above, it should not be all that hard for a strict impartialist to do what he thinks he has most reason to do. Of course, once the torture begins, he will presumably feel an irresistible urge to swap places with the other person; showing up voluntarily for torture every day for 50 years would certainly exceed human motivational capacities. But as the case is formulated, he just has to utter a simple "yes" and then the demon will take care of the rest. Would Nagel, in a case like this, be overpowered with self-love or a fear of pain in the distant future, so that he just couldn't bring himself to volunteer for the torture even if that is what he thinks there is by far most reason to

¹²⁷ Utilitarians often appeal to the psychological limits of altruistic motivation when arguing that their doctrine does not make as extreme demands as one might think (see e.g. Shaw 1999, 262).

do? I cannot help thinking that Nagel would say no, not because of weakness of will, but because of an overwhelming appearance that there is an authoritative *reason* to do so.

Nagel questioned whether there really are any egoists. I would like to question whether there really are any strict impartialists. More generally, I suspect that philosophers often do not believe quite what they say they believe about their reasons concerning other people. Of course, I am not saying that that this apparent insincerity constitutes any kind of proof that any kind of impartialism is mistaken. It just annoys me.

9.4 Final remarks

In this chapter, I have reviewed several ways in which my version of hedonistic egoism is counterintuitive. As a summary, we might put them all together in a single case. Assume that you are faced with the following to options:

(1) You and your child go on to lead long lives full of pleasure, knowledge, virtue and all other things that philosophers have thought worth having for its own sake.

(2) Your child will be miserable for the rest of his life, while you will be connected to an experience machine, experiencing slight sadistic pleasure for a million years, so that the sum of hedonic value in your life would be higher than in (1) by a tiny amount.

According to the theory I have been defending in this thesis, (2) is a more choice-worthy option than (1). It might seem as if no argument could possibly get us to accept a conclusion as violently counterintuitive as this, and in particular not an argument of the fairly sweeping and speculative character given in chapters 7-8. However, the idea is not that my argument for hedonistic egoism is supposed to be so strong that we can live with some highly counterintuitive implications about particular cases. Rather, the point is that if my argument is sound, intuitions such as those elicited by the choice between (1) and (2) do not reflect normative truths at all, and so do not in any way count for or against my view. No amount of intuitive horror of this kind should get us to change our minds about whether the outlook on reasons presented in this thesis is plausible.

It is certainly possible to question my argument in other ways. Perhaps I have given the wrong account of pleasure and pain, misconstrued the empirical findings on the role of affect in reasoning, or made mistakes in the criticisms of reasons pluralism and the other alternative views. Or perhaps the alternatives can be improved so that the criticisms no longer apply. But

in order to disbelieve hedonistic egoism, one would have to actually show that the argument is flawed in at least one of these senses. Simply thinking that it must be since a theory like hedonistic egoism could not possibly be correct would be sheer dogmatism.

Bibliography

- Allen, Colin; Fuchs, Perry N.; Shriver, Adam & Wilson, Hilary D., “Deciphering Animal Pain”, in Aydede, Murat (ed.), *Pain* (MIT Press, 2005)
- Alston, William, “Pleasure”, in Edwards, Paul (ed.), *Encyclopedia of Philosophy* (Macmillan, 1967)
- Annas, Julia, “Virtue Ethics and the Charge of Egoism”, in Bloomfield, Paul (ed.), *Morality and Self-Interest* (Oxford University Press, 2008)
- Ariely, Dan, *Predictably Irrational* (Harper Collins, 2008)
- Aristotle, *Nicomachean Ethics* (Dover Publications, 1998)
- Audi, Robert, *Moral Knowledge and Ethical Character* (Oxford University Press, 1997)
- Aydede, Murat, “An Analysis of Pleasure Vis-à-Vis Pain”, *Philosophy and Phenomenological Research* 69 (2000)
- Baier, Kurt, *The Moral Point of View* (Cornell University Press, 1958)
- Baier, Kurt, “Egoism”, in Singer, Peter (ed.), *A Companion to Ethics* (Blackwell, 1993)
- Baker, Alan, “Simplicity”, in Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Stanford University, 2010)
- Bentham, Jeremy, *An Introduction to the Principles of Morals and Legislation* (Kessinger, 2005, reprint of 1823 edition)
- Berridge, Kent C., “Pleasures of the Brain”, *Brain & Cognition* 52 (2003)
- Berridge, Kent C., “Motivation concepts in behavioral neuroscience”, *Physiology & Behavior* 81 (2004)
- Beyleveld, Deryck, *The Dialectical Necessity of Morality* (The University of Chicago Press, 1991)
- Blackburn, Simon, *Essays in Quasi-Realism* (Oxford University Press, 1993)
- Blackburn, Simon, *Ruling Passions* (Oxford University Press, 1998)
- Boyd, Richard N., “How to Be a Moral Realist”, in Sayre-McCord, Geoffrey (ed.), *Essays on Moral Realism* (Cornell University Press, 1988)
- Brandt, Richard B., *A Theory of the Good and the Right* (Oxford University Press, 1979)
- Brax, David, *Hedonism as the Explanation of Value* (PhD thesis, University of Lund, 2009)
- Brink, David O., *Moral Realism and the Foundations of Ethics* (Cambridge University Press, 1989)
- Brink, David O., “A Reasonable Morality”, *Ethics* 104 (1994)

- Broad, C. D., *Five Types of Ethical Theory* (Routledge, 2000, first published in 1930)
- Broome, John, "More pain or less?", *Analysis* 56 (1996)
- Broome, John, "Normative Requirements", *Ratio* 12 (1999)
- Broome, John, "Reasons", in Wallace, Jay R; Pettit, Philip; Scheffler, Samuel & Smith, Michael (ed.), *Reason and Value* (Oxford University Press, 2004)
- Brunton, J. A., "Egoism and Morality", *The Philosophical Quarterly* 6 (1956)
- Butler, Joseph, "Upon the Love of Our Neighbour", first published in 1726, reprinted in Milo, Ronald D. (ed.), *Egoism and Altruism* (Wadsworth Publishing Company, 1973)
- Casebeer, William D., "Processes and Moral Emotions", in Sinnott-Armstrong, Walter (ed.), *Moral Psychology* (The MIT Press, 2008)
- Chang, Ruth, "Can Desires Provide Reasons for Action?", in Wallace, Jay R; Pettit, Philip; Scheffler, Samuel & Smith, Michael (ed.), *Reason and Value* (Oxford University Press, 2004)
- Clark, Austen, "Painfulness is not a quale", in Aydede, Murat (ed.), *Pain* (MIT Press, 2005)
- Clore, G. L. & Huntsinger, J. R. "How emotions inform judgment and regulate thought", *Trends in Cognitive Science* 11, (2007)
- Cocking, Dean & Oakley, Justin, "Indirect Consequentialism, Friendship, and the Problem of Alienation", *Ethics* 106 (1995)
- Crisp, Roger, "Value, reasons and the structure of justification: How to Avoid Passing the Buck", *Analysis* 65 (2005)
- Crisp, Roger, *Reasons and the Good* (Oxford University Press, 2006)
- Crisp, Roger, "Sidgwick's Hedonism", in Bucolo, Placido; Crisp, Roger & Schultz, Bart (ed.), *Proceedings of the World Congress on Henry Sidgwick* (Universita degli Studi di Catania, 2007)
- Cuneo, Terence, *The Normative Web* (Oxford University Press, 2007)
- Dancy, Jonathan, "Enticing Reasons", in Wallace, Jay R; Pettit, Philip; Scheffler, Samuel & Smith, Michael (ed.), *Reason and Value* (Oxford University Press, 2004)
- Daniels, Norman: "Wide Reflective Equilibrium and Theory Acceptance in Ethics", *Journal of Philosophy* 76 (1979)
- Darwall, Stephen, "Human Morality's Authority", *Philosophy and Phenomenological Research* 55 (1995)
- Darwall, Stephen, *The Second-Person Standpoint* (Harvard University Press, 2006)
- Duncker, Karl, "On Pleasure, Emotion, and Striving", *Philosophy and Phenomenological Research* 1 (1941)

- Epstein, Seymour, "Integration of the cognitive and the psychodynamic unconscious", *American Psychologist* 49 (1994)
- Feinberg, Joel, "Psychological Egoism", in Feinberg, Joel & Shafer-Landau, Russ (ed.), *Reason and Responsibility* (Thomson Wadsworth, 2005)
- Feldman, Fred, "Two Questions about Pleasure", in Austin, David F. (ed.), *Philosophical Analysis: A Defense by Example* (Kluwer Academic Publishers, 1988)
- Feldman, Fred, "On the Intrinsic Value of Pleasures", *Ethics* 107 (1997)
- Feldman, Fred, "Hedonism", in Becker, Lawrence C. & Becker, Charlotte B. (ed.), *Encyclopedia of Ethics* (Routledge, 2001)
- Feldman, Fred, *Pleasure and the Good Life* (Oxford University Press, 2004)
- Finlay, Stephen, "Too Much Morality", in Bloomfield, Paul (ed.), *Morality and Self-Interest* (Oxford University Press, 2008)
- Frankena, William K., *Ethics*, Second Edition (Prentice Hall, 1973)
- Garner, Richard T., "On the Genuine Queerness of Moral Properties and Facts", *Philosophy* 68 (1990)
- Gauthier, David, *Morals by Agreement* (Oxford University Press, 1986)
- Gert, Joshua, "Korsgaard's Private-Reasons Argument", *Philosophy and Phenomenological Research* 64 (2002)
- Gewirth, Alan, *Reason and Morality* (The University of Chicago Press, 1978)
- Gewirth, Alan, "Reply to My Critics", in Regis Jr., Edward (ed.), *Gewirth's Ethical Rationalism* (The University of Chicago Press, 1984)
- Gibbard, Allan, "The Reasons of a Living Being", *Proceedings and Addresses of the American Philosophical Association* 76 (2002)
- Gigerenzer, Gerd, "Moral Intuitions = Fast and Frugal Heuristics?", in Sinnott-Armstrong, Walter (ed.), *Moral Psychology* (The MIT Press, 2008)
- Goldman, Alan H., "The Case Against Objective Values", *Ethical Theory and Moral Practice* 11 (2008)
- Goldstein, Irwin, "Why People Prefer Pleasure to Pain", *Philosophy* 55 (1980)
- Gowans, Chris, "Moral Relativism", in Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Stanford University, 2010)
- Grahek, Nikola, *Feeling Pain and Being in Pain* (MIT Press, 2001)
- Greene, Joshua, "The Secret Joke of Kant's Soul" in Sinnott-Armstrong, Walter (ed.), *Moral Psychology* (The MIT Press, 2008)
- Griffin, James, *Well-being* (Oxford University Press, 1986)

- Haidt, J., “The emotional dog and its rational tail: A social intuitionist approach to moral judgment”, *Psychological Review* 108 (2001)
- Hampton, Jane E., *The Authority of Reason* (Cambridge University Press, 1998)
- Hardin, Garrett, “The Tragedy of the Commons”, *Science* 162 (1968)
- Hare, R. M., *Moral Thinking* (Oxford University Press, 1981)
- Hare, R. M., “Do Agents Have to be Moralists?”, in Regis Jr., Edward (ed.), *Gewirth’s Ethical Rationalism* (The University of Chicago Press, 1984)
- Hare, R. M., *Essays in Ethical Theory* (Oxford University Press, 1989)
- Harman, Gilbert: *The Nature of Morality* (Oxford University Press, 1977)
- Harman, Gilbert, “Moral Relativism”, in Harman, Gilbert & Thompson, Judith Jarvis, *Moral Relativism and Moral Objectivity* (Blackwell, 1996)
- Haslett, D. W., *Capitalism with Morality* (Oxford University Press, 1994)
- Heathwood, Chris, “The Reduction of Sensory Pleasure to Desire”, *Philosophical Studies* 133 (2007)
- Henden, Edmund, “Restrictive Consequentialism and Real Friendship”, *Ratio* 20 (2007)
- Hewitt, Sharon, “What Do Our Intuitions About the Experience Machine Really Tell Us About Hedonism?”, *Philosophical Studies* (forthcoming)
- Hills, Allison, “Value, Reason and Hedonism”, *Utilitas* 20 (2008)
- Hobbes, Thomas, *Leviathan* (Cambridge University Press, 1996, first published in 1651)
- de Houwer, Jan, Thomas, Sarah & Bayens, Frank, “Associative learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning”, *Psychological Bulletin* 127 (2001)
- Hsieh, Nien-hê, “Incommensurable Values”, in Zalta, Edward N. (ed.), *Stanford Encyclopedia of Philosophy*, (Stanford University, 2010)
- Hume, David, “The Sceptic”, first published in 1743, reprinted in Cahn, Steven M. & Vitriano, Christine (ed.), *Happiness* (Oxford University Press, 2008)
- Hume, David, *An Enquiry Concerning the Principles of Morals* (Hackett Publishing Company 1983, first published in 1751)
- Hunt, Lester, “Flourishing Egoism”, *Social Philosophy and Policy* 16 (1999)
- Hynes, Catherine A., “Morality, Inhibition, and Propositional Content”, in Sinnott-Armstrong, Walter (ed.), *Moral Psychology* (The MIT Press, 2008)
- van Ingen, John, *Why Be Moral?* (Peter Lang Publishing, 1994)

- Jones, C. R., Fazio, R. H., & Olson, M. A., “Implicit misattribution as a mechanism underlying evaluative conditioning”, *Journal of Personality and Social Psychology* 96 (2009)
- Joyce, Richard: *The Myth of Morality* (Cambridge University Press, 2001)
- Joyce, Richard, *The Evolution of Morality* (The MIT Press, 2006)
- Kahane, Guy, “Pain, Dislike and Experience”, *Utilitas* 21 (2009)
- Kalin, Jesse, “In Defense of Egoism”, in Gauthier, David P. (ed.), *Morality and Rational Self-interest* (Prentice-Hall, 1970)
- Kalin, Jesse, “The Persistence of Egoism”, in Regis Jr., Edward (ed.), *Gewirth’s Ethical Rationalism* (The University of Chicago Press, 1984)
- Kant, Immanuel, *Groundwork of the Metaphysics of Morals* (Cambridge University Press, 1997, first published in 1785)
- Kapur, Neera Badhwar, “Why It Is Wrong to Be Always Guided by the Best: Consequentialism and Friendship”, *Ethics* 101 (1991)
- Katz, Leonard, “Hedonic Reasons as Ultimately Justifying and the Relevance of Neuroscience”, in Sinnott-Armstrong, Walter (ed.), *Moral Psychology* (MIT Press, 2008)
- Katz, Leonard, “Pleasure”, in Zalta, Edward N. (ed.), *Stanford Encyclopedia of Philosophy*, (Stanford University, 2010)
- Kavka, Gregory S., “The Reconciliation Project”, in Zimmermann, David & Copp, David (ed.), *Morality, Reason and Truth* (Rowman and Allanheld, 1985)
- Kawall, Jason, “The Experience Machine and Mental State Theories of Wellbeing”, *The Journal of Value Inquiry* 33 (1999)
- Kekes, John, “Happiness”, *Mind* 91 (1982)
- Kolodny, Niko, “Why Be Rational?”, *Mind* 114 (2005)
- Korsgaard, Christine, “Two Distinctions in Goodness”, *Philosophical Review* 92 (1983)
- Korsgaard, Christine, *The Sources of Normativity* (Cambridge University Press, 1996)
- Kraut, Richard, “Desire and the Human Good”, *Proceedings and Addresses of the American Philosophical Association* 68 (1994)
- Kringelbach, Morten, “The Human Orbitofrontal Cortex: Linking Reward to Hedonic Experience”, *Nature Reviews Neuroscience* 6 (2005)
- Layard, Richard, *Happiness* (Penguin Books, 2005)
- Leiter, Brian, *Nietzsche on Morality* (Routledge 2002)
- Leknes, Siri & Tracey, Irene, “A common neurobiology for pain and pleasure”, *Nature Reviews Neuroscience* 9 (2008)

- Lemos, Noah, “High-Minded Egoism and the Problem of Priggishness”, *Mind* 93 (1984)
- Lewis, Clarence Irving, *An Analysis of Knowledge and Valuation* (The Open Court Publishing Company, 1946)
- Lewis, David, “Survival and identity”, in Rorty, Amélie O. (ed.), *The Identities of Persons* (University of California Press, 1976)
- Lloyd Thomas, D. A., “Why Should I Be Moral?”, *Philosophy* 45 (1970)
- Loewenstein, G. F.; Weber, E. U.; Hsee, C. K. & Welch, E. S., “Risk as feelings”, *Psychological Bulletin* 127 (2001)
- Mackie, John, *Ethics* (Pelican Books, 1977)
- Mason, Elinor, “Can an Indirect Consequentialist Be a Real Friend?”, *Ethics* 108 (1998)
- Mason, Elinor, “The Nature of Pleasure: A Critique of Feldman”, *Utilitas* 19 (2007)
- Matson, Wallace I., “Hegesias the Death-Persuader; Or, the Gloominess of Hedonism”, *Philosophy* 73 (1998)
- Medlin, Brian, “Ultimate Principles and Ethical Egoism”, *Australasian Journal of Philosophy* 35 (1957)
- Mendola, Joseph, “Objective Value and Subjective States”, *Philosophy and Phenomenological Research* 50 (1990)
- Mendola, Joseph, “Intuitive Hedonism”, *Philosophical Studies* 128 (2006)
- Mill, John Stuart, *Utilitarianism* (Broadview Press, 2000, reprint of the 1871 edition)
- Moore, Andrew, “Hedonism”, in Zalta, Edward N. (ed.), *Stanford Encyclopedia of Philosophy*, (Stanford University, 2010)
- Moore, G. E., *Principia Ethica* (Cambridge: Cambridge University Press, 1993, reprint of the 1903 edition)
- Nagel, Thomas, *The Possibility of Altruism* (Princeton University Press, 1970)
- Nagel, Thomas, *The View from Nowhere* (Oxford University Press, 1986)
- Nietzsche, Friedrich, *Beyond Good and Evil* (Dover Publications, 1997, first published in 1886)
- Nietzsche, Friedrich, *On the Genealogy of Morality* (Cambridge University Press, 1994, first published in 1887)
- Nisbett, Richard E. & Wilson, Timothy, “Telling More Than We Can Know: Verbal Reports on Mental Processes”, *Psychological Review* 84 (1977)
- Nozick, Robert, *Anarchy, State, and Utopia* (Blackwell, 1974)
- Nozick, Robert, *The Examined Life: Philosophical Meditations* (Simon & Schuster, 1990)

- Olson, Jonas, "Buck-Passing and the Wrong Kinds of Reasons", *The Philosophical Quarterly* 54 (2004)
- Olson, Robert G., "Ethical Egoism and Social Welfare", *Philosophy and Phenomenological Research* 21 (1961)
- Österberg, Jan, *Self and Others* (Kluwer Academic Publishers, 1988)
- Panksepp, Jaak, "On the Neuro-Evolutionary Nature of Social Pain, Support, and Empathy", in Aydede, Murat (ed.), *Pain* (MIT Press, 2005)
- Parfit, Derek, *Reasons and Persons* (Oxford University Press, 1984)
- Parfit, Derek, *On What Matters* (Oxford University Press, forthcoming)
- Plato, *The Republic* (Oxford University Press, 1993)
- Prinz, Jesse J., "Is Morality Innate?", in Sinnott-Armstrong, Walter (ed.), *Moral Psychology* (MIT Press, 2008)
- Rachels, Jim, "Two Arguments against Ethical Egoism", *Philosophia* 4 (1974)
- Rachels, Jim, *The Elements of Moral Philosophy*, 4th edition (McGraw-Hill, 2002)
- Rachels, Stuart, "Is Unpleasantness Intrinsic to Unpleasant Experiences?", *Philosophical Studies* 99 (2000)
- Rachels, Stuart, "Nagelian Arguments against Egoism", *Australasian Journal of Philosophy* 80 (2002)
- Rachels, Stuart, "A Defense of Two Optimistic Claims in Ethical Theory", *Philosophical Studies* 112 (2003)
- Rachels, Stuart, "Six Theses of Pleasure", *Philosophical Perspectives* 18 (2004)
- Rachels, Stuart, "On Three Alleged Theories of Rational Behavior", *Utilitas* 21 (2009)
- Railton, Peter, "Alienation, Consequentialism, and the Demands of Morality", *Philosophy and Public Affairs* 13 (1984)
- Railton, Peter, "Moral Realism", *The Philosophical Review* 95 (1986)
- Rand, Ayn, *The Virtue of Selfishness* (Signet, 1964)
- Rawls, John, *A Theory of Justice*, Revised Edition (Oxford University Press, 1999a)
- Rawls, John: *Collected Papers* (Harvard University Press, 1999b)
- Rawls, John, *Justice as Fairness: A Restatement* (Harvard University Press, 2001)
- Raz, Joseph, *Engaging Reason* (Oxford University Press, 1999)
- Regan, Donald H., "How to be a Moorean", *Ethics* 113 (2003)
- Regis Jr., Edward, "What is Ethical Egoism?", *Ethics* 91 (1980)
- Risse, Mathias, "Nietzsche on Selfishness, Justice, and the Duties of the Higher Men", in Bloomfield, Paul (ed.), *Morality and Self-Interest* (Oxford University Press, 2008)

- Ross, David, *The Right and the Good* (Oxford University Press, 1930)
- Rottenstreich, Y. & Hsee, C. K., “Money, kisses, and electric shocks: An affective psychology of risk”, *Psychological Science* 12 (2001).
- Russell, James A., “Core Affect and the Psychological Construction of Emotion”, *Psychological Review* 110 (2003)
- Ryberg, Jesper; Tännsjö, Torbjörn & Arrhenius, Gustav, “The Repugnant Conclusion”, in Zalta, Edward N. (ed.), *Stanford Encyclopedia of Philosophy*, (Stanford University, 2010)
- Scanlon, Thomas, *What We Owe to Each Other* (Harvard University Press, 1998)
- Scanlon, Thomas: “Rawls on Justification”, in Freeman, Samuel (ed.), *Cambridge Companion to Rawls* (Cambridge University Press, 2003)
- Scheffler, Samuel, *Human Morality* (Oxford University Press, 1992)
- Scheffler, Samuel, “Reply to Three Commentators”, *Philosophy and Phenomenological Research* 55 (1995)
- Scheffler, Samuel, “Relationships and Responsibilities”, *Philosophy and Public Affairs* 29 (1997)
- Schlick, Moritz, *Problems of Ethics* (Dover Publications, 1939)
- Schwarz, Norbert & Clore, Gerald L., “Mood, misattribution, and judgments of well-being”, *Journal of Personality and Social Psychology* 45 (1983)
- Seneca, Lucius Annasus, *Moral Essays*, vol. 2 (W. Heinemann, 1932)
- Shafer-Landau, Russ, *Moral Realism* (Oxford University Press, 2003)
- Shafer-Landau, Russ, “Defending Ethical Intuitionism”, in Sinnott-Armstrong, Walter (ed.), *Moral Psychology* (MIT Press, 2008)
- Shaver, Robert, *Rational Egoism* (Cambridge University Press, 1999)
- Shaver, Robert, “Egoism”, in Zalta, Edward N. (ed.), *Stanford Encyclopedia of Philosophy*, (Stanford University, 2010)
- Shaw, William H., *Contemporary Ethics* (Blackwell, 1999)
- Sidgwick, Henry, *The Methods of Ethics*, Seventh Edition (Hackett Publishing Company, 1981, first published in 1907)
- Silverstein, Matthew, “In Defense of Happiness: A Response to the Experience Machine”, *Social Theory and Practice* 26 (2000)
- Singer, Peter, *Practical Ethics*, Second Edition (Cambridge University Press, 1993)
- Singer, Peter, “Ethics and Intuitions”, *The Journal of Ethics* 9 (2005)
- Sinnott-Armstrong, Walter, “Framing Moral Intuitions”, in Sinnott-Armstrong, Walter (ed.), *Moral Psychology* (MIT Press, 2008)

- Slote, Michael A., *Beyond Optimizing* (Harvard University Press, 1989)
- Slovic, Paul, "The Affect Heuristic" in Gilovich, Thomas; Griffin, Dale & Kahneman, Daniel (ed.) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge University Press, 2002)
- Smith, Michael, *The Moral Problem* (Blackwell Publishing, 1994)
- Smith, Tara, *Viable Values* (Rowman & Littlefield, 2000)
- Smith, Tara, *Ayn Rand's Normative Ethics* (Cambridge University Press, 2006)
- Smuts, Aaron, "The Feels Good Theory of Pleasure", *Philosophical Studies* (forthcoming)
- Sobel, David, "Pain for Objectivists: The Case of Matters of Mere Taste", *Ethical Theory and Moral Practice* 8 (2005)
- Southwood, Nicholas, "Vindicating the Normativity of Rationality", *Ethics* 119 (2007)
- Sprigge, T. L. S., *The Rational Foundations of Ethics* (London: Routledge and Kegan Paul, 1988)
- Stirner, Max, *The Ego and His Own* (Dover Publications, 2005, first published in 1845)
- Stocker, Michael, "The Schizophrenia of Modern Ethical Theories", *The Journal of Philosophy* 73 (1976)
- Street, Sharon, "A Darwinian Dilemma for Realist Theories of Value", *Philosophical Studies* 127 (2006)
- Stroud, Sarah, "Moral Overridingness and Moral Theory", *Pacific Philosophical Quarterly* 79 (1998)
- Sturgeon, Nicholas L., "Altruism, Solipsism, and the Objectivity of Reasons", *The Philosophical Review* 83 (1974)
- Sturgeon, Nicholas L., "Moral Explanations", in Sayre-Mccord, Geoffrey (ed.), *Essays on Moral Realism* (Cornell University Press, 1988)
- Sumner, L. W., *Welfare, Happiness, and Ethics* (Oxford University Press, 1996)
- Swenson, Adam, "Pain's Evils", *Utilitas* 21 (2009)
- Tännsjö, Torbjörn, *Moral Realism* (Rowman & Littlefield, 1990)
- Tännsjö, Torbjörn, *Hedonistic Utilitarianism* (Edinburgh University Press, 1998)
- Tännsjö, Torbjörn, "Why We Ought to Accept the Repugnant Conclusion", *Utilitas* 14 (2002)
- Tännsjö, Torbjörn, "Narrow Hedonism", *Journal of Happiness Studies* 7 (2007)
- Tännsjö, Torbjörn, *From Reasons to Norms* (Springer, 2010)
- Taylor, Richard, *Virtue Ethics* (Prometheus Books, 2002)
- Timmermann, Jens, "Too Much of a Good Thing? Another Paradox of Hedonism", *Analysis* 65 (2005)

- Tye, Michael, "Another Look at Representationalism about Pain", in Aydede, Murat (ed.), *Pain* (MIT Press, 2005)
- Velleman, J. David, "Well-being and Time", *Pacific Quarterly* 72 (1991)
- Whiteley, C. H., "Morality and Egoism", *Mind* 85 (1976)
- Wilcox, William H., "Egoists, Consequentialists, and Their Friends", *Philosophy and Public Affairs* 16 (1987)
- Williams, Bernard, "Internal and External Reasons", in Harrison, Ross (ed.), *Rational Action* (Cambridge University Press, 1979)
- Williams, Bernard, *Ethics and the Limits of Philosophy* (Fontana Press, 1985)
- Williams, Gardner, "Individual, Social, and Universal Ethics", *The Journal of Philosophy* 45 (1948)
- Wood, Allen W., *Kant's Ethical Thought* (Cambridge University Press, 1999)
- Wood, Allen W., *Kantian Ethics* (Cambridge University Press, 2008)
- World Health Organization, *Investing in Mental Health* (2003, available on http://www.who.int/mental_health/en/investing_in_mnh_final.pdf)
- von Wright, G. H., *The Varieties of Goodness* (Routledge & Kegan Paul, 1963)
- Zimmermann, Michael J., "On the Intrinsic Value of States of Pleasure", *Philosophy and Phenomenological Research* 41 (1980)