Hedonism as the Explanation of Value

Brax, David

2009

Hedonism as the Explanation of Value

Hedonism as the Explanation of Value

David Brax

David Brax (2009)
Hedonism as the Explanation of Value

ISBN 978-91-628-7855-9

Printed at Media-Tryck Sociologen, 2009
Cover illustration by Johanna Kindvall

For Alice

Table of Contents

Preface

It started with a rather simple idea, set to solve a particular problem in the theory of value. Well, actually, there were two problems: the first was to find a plausible version of preferentialism, i.e. of the view that what is valuable depends on preferences, while the other was to make sense of how a value that depends on preferences might still be *intrinsic* to what is valuable. The problem, in short, is that if the value of things depends on our preferences, it seems to depend on features that are *extrinsic* to it. Rather than resolving the issue by abandoning the notion of intrinsic value, a move that was very much in style at the time, I set about developing a notion of preference-dependent value that was compatible with it. The reason for this, however, was not any theoretical attachment to intrinsic value, but rather than none of the examples of non-intrinsic so called "final" values struck me as very persuasive. The problem with most versions of preferentialism that I came across was not that they violated the, let's face it, rather academic notion of intrinsic value, but that they seemed to get the relation between our preferences and the valuable state wrong.

A problem facing preference-based theories of value, be they substantial claims about what's good, or meta-ethical claims about the nature of value, is the existence of irrational, misdirected preferences, which fail to target things that would be good for us. The solution often offered is that the preferences relevant to what's good are those that are *ideal*: i.e. the preference we *would* have if, say, we were fully rational, fully informed, freed of cognitive infirmities. Again, this solution struck me as unsatisfactory, as missing the point of what's plausible in a preference-oriented theory about what makes something good.

In what I later realised was a patently Epicurean move, I believed that the solution to the problem of misdirected, irrational preferences was to make the relation between preferences and value much *closer*. The only preferences that track value are those that take as their object our own *experiences*. We can always be, and often are, mistaken about the nature and importance of external facts, but we seem to have a privileged access to our own experiences. This ensures that we know what we commit ourselves to when we declare our preference for them. And yet the relation did not strike me as being quite close enough: it seemed insufficient to say that our preferences took those valuable states as their *objects* when it was so obvious that what *made* those objects valuable was the relation to that preference. The point of preferentialism, I took it, was precisely that the objects of preferences would have no value if they occurred on their own: the preferences did not pick out a value property that was there in the object already. Preferences and experiences both being mental states, it struck me that the valuable experiences where partly *constituted* by the preference, that the relation between them was not merely formal, but concrete and interactive.

The resulting mental states, quite clearly, were *pleasures* and the theory of value I defended, consequentially, a version of hedonism. This theory rather elegantly, as I thought (being 22 years old at the time), combined a plausible theory of pleasure with a preference-oriented view about value, compatible with the notion of intrinsic value.

Then something happened. Autobiographically, I guess one could say that cognitive science happened, which caused the realisation that I really didn't know enough about pleasure. What *is* pleasure? And how does it relate to motivation, evaluation and action? What role does it play in human psychology? Seeing how hedonists used to be very engaged with scientific psychology, and that the notion of pleasure I had in mind suggested a concrete relation between preferences and pleasures, surely I would have to look into this matter too. That this angle of hedonism had been neglected for so long struck

me as something of an outrage. That is, until someone brought to my attention a dissertation written in the mid-eighties by one Leonard D. Katz called "Hedonism as the metaphysics of mind and value". By this time, I'd started work on my own dissertation, and reading Katz's book made my heart sink. Here was, in an eerie, uncanny way, the very book I wanted to write. In fact, the book I was already engaged in writing. It defended a notion of pleasure very close to my own, and it did so on a very ambitious basis of philosophical reasoning, extensive reading of historical texts and a great deal of psychological science. For a while, the only thing that made me believe that there might be a point in my continuing writing at all was the fact that almost twenty years had gone by, and things had happened in affective neuroscience. I met Dr Katz in a bookstore in Boston in September 2006, after engaging in a very encouraging correspondence. He had then recently published what is, and will for a long time continue to be, the best survey of philosophical and, arguably, scientific theories about pleasure. In the conversation, I mentioned my qualms about writing on the same subject and with a very similar approach, but he reassured me that our views where sufficiently different and mine sufficiently independent for me not to worry. Besides, it is hardly surprising that we would have come up with the same idea since it is, roughly, *true*.

During the same trip to Boston, I also visited Fred Feldman, a philosopher whose work on pleasure was the main inspiration for my taking up the subject in the first place. It was his writing about the problem to square a preference-based theory of pleasure with the notion of intrinsic value that made me develop my own view. Our solutions to the problem are, in one sense, very similar but our theoretical approaches are very different. Both these facts make the differences illuminating.

My Ph-D position was brought about in September 2003 under the project "Philosophical Theories About Value", financed by the Bank of Sweden Tercentenary Foundation, which included my supervisors Wlodek Rabinowicz and Toni Rønnow-Rasmussen. My interest in intrinsic value, and the finer

points about the ontological classification of value-bearers soon gave way to more general questions about the nature of value. Whereas I started out more or less assuming an unproblematic notion of intrinsic value, I became interested in what this thing actually was, and how to make sense of it. Having just previously spent six months on a paper on the nature of consciousness, I noticed a striking similarity between the problem of value and the problem of subjective experience: both tend to resist reduction in naturalistic, functional terms. Perhaps, I thought, they are at least partly the *same* problem.

Hedonism is a controversial position. It seems to go against many of our dearly held beliefs about what is good in life. Hedonists have generally tried to get around this problem by *explaining* such beliefs *away*. Pleasure, they, *we,* claim is the only thing that *really* has value. This, I figured, is not merely an act of self-defence on behalf of the hedonist, but actually *essential* to the type of theory of value the hedonist should be defending. Hedonism is best understood as an *explanatory* approach to value: pleasure is a plausible candidate as the only good because pleasure is involved in the best explanation of our evaluative behaviours and experiences. This approach to value is already part of the empirical interest in the nature of pleasure, and its function in human psychology.

I spent four very inspiring and exhausting months in Oxford in the spring of 2007, under the occasional supervision of Dr Krister Bykvist. While there, I had tea, I talked to people, I attended lectures and workshops and a high-table dinner. I got engaged in a punt on the river Cherwell. I also made the tactical blunder to find yet another approach to hedonism, based in meta-ethical *naturalism*. Naturalism is, I believe, the best approach for an ambitious hedonist, and a naturalist, explanatory, empirically informed approach to value supports a version of hedonism. This claim, I suppose, makes up much of what is original in this book. Taking this road was a tactical blunder insofar as I've spent, as anyone who knows me and has had to put up with me will tell you, far too much time trying to get this bit right.

This book could have been ten times the size it is. My aim was to find out the truth about hedonism and this project proved to be almost impossibly inclusive. It concerns the philosophy of mind and value, but also the cognitive and affective sciences, philosophy of language and science, the nature of theory and explanation, even metaphysics. As it stands, then, the book is lacking in many respects. Possibly, I should have focused on an even smaller portion of the project, but I simply couldn't bring myself to do so.

There are two things of note that I ended up *not* doing. The first is Exegesis. You will find very little discussion of the literature here. For the most part, my M.O. is quite straight-forward reasoning, and my primary concern is to develop a fairly original view of my own. Quotes and borrowed arguments are inserted mainly to bring the reasoning forward and for the sake of illustration. I apologize if this means that I make some faulty interpretations along the way. This also means that the book is not very polemical in its structure. The main point of it is a positive argument for a theory of pleasure and value. It suggests an approach to these matters that seems to me interesting and true.

The other absentee concerns the science. I've spent a fair amount of time reading up on the affective science literature. Insofar as I am any judge, the findings in this discipline so far are consistent with, and even support, my views. But I'm not an expert in this field. For this reason, I've hesitated to include references to this literature in the text. I ended up including a relatively small amount of text addressing the scientific research directly: Mostly, what I write about it is a call for philosophers, like myself, to pay more attention to this research. If we don't, we risk making unfounded assumptions, and develop theories based on what we take to be "common sense", which, it turns out, is far from how things actually work. If I had decided to include a review of this research in the book, it would have been a very selective one, and I lack the right background to write such a review in the proper context. This decision is the only display of modesty you'll find in these pages.

In addition to the financing of my PhD-position by the Bank of Sweden Tercentenary Foundation, I've received financial support from STINT, Erik och Gurli Hultengrens fond för filosofi and Stiftelsen Fil dr. Uno Otterstedts fond for främjande av vetenskaplig undervisning och forskning. For this I'm very grateful.

Thanks are due to the people who have listened and commented to the talks I've delivered at conferences and seminars, in particular to the participants of the higher seminar at the philosophy departments at Lund University, Stockholm University, Uppsala University, the Royal Institute of Technology and the James Martin seminar at the philosophy department at Oxford University. Thanks are also due to the students who took my course in Hedonism during the spring of 2004 and autumn of 2007. My stay in Oxford was made particularly pleasant by Jens Johansson, Jonas Olson, Brian McElwee and Roger Crisp, as well as by the UEHIRO center for practical ethics, which very kindly cleared a desk for me. At the philosophy department in Lund, my roommates provided my main source of psychological support; Petra Björne and Tomas Persson, both now accomplished PhD's. From a distance, Marie Lundstedt provided the same service. My supervisors Wlodek Rabinowicz and Toni Rønnow-Rasmussen have very patiently overseen my dithering between theoretical options, and provided excellent advice on how to move forward. It is not their fault that I've not always taken that advice to heart. Above all I thank my wife Alice, whose patience, humour, love and support over the last five years is by far the most interesting and wonderful thing I've come across during my research. I love you.

What I'm thanking for is patience, basically. The Swedish word for 'patience' is 'tålamod', a word that makes a point about patience being a kind of courage. My gratefulness for the patience shown by those concerned is for their courage in letting me keep working on this project for all this time.

Stockholm 20/7 2009

Part 1: Pleasure

1.1 What is Pleasure?

1.1.1 Introduction

Pleasure is of the utmost importance. This is the guiding principle behind all that follows. Pleasure is central to all sentient life; it is central to emotion, it plays a pivotal role in action, in decision, in motivation and it is absolutely central to what's good in life. Indeed, the suggestion put forward in this book is that pleasure *is* the good. The argument for that thesis is primarily confined to part 2. This part, for which I presume there is independent interest, is concerned with what pleasure is. This project is indispensable for a hedonistic theory of the good, since we need to know what it is that the hedonist claim is good. Luckily, the most plausible account of pleasure as such fits very well with the account of value that I have in store.

The question "what is pleasure" should meet with an immediate first qualification: what *kind* of a thing is pleasure? A natural suggestion is that pleasure is a kind of experience. Experiences are regularly distinguished by how they *feel*, so pleasure would then presumably be a class of experiences distinct by their felt quality. That is, what *makes* these experiences pleasures is how they feel. This is arguably the historically dominating view of pleasure, but it has received a lot of criticism. If not an experience itself, pleasure is at the very least something that can be *experienced:* it might be the *content* or *object* of an experience. It might belong to the more general genus of *mental states*. Mental states in general can be distinguished not only by how they feel but by their content or by their function, so if the distinctive feeling view fails, there are other options. While still being experiences, pleasures would then be distinguished, not by intrinsic, but by relational properties: an experience or a mental state is a pleasure if and only if it stands in some relation to some attitude that the agent has, say. This has become the majority view among

philosophers writing extensively on pleasure, at least since Henry Sidgwick's The Methods of Ethics.[1]

A further option along these lines is to say that, as a mental state, pleasure itself might be intentional, i.e. not the *object* of an attitude, but an attitude in its own right. It could then be distinguished by the kind of object it takes, or by the operation it performs on that object. Pleasure could be understood as a belief or judgment with some particular content, or as the representation of some particular content. The critical element might be an attitude like "Taking pleasure in"[2], or enjoying[3], in which case there are questions to answer about what kind of object the attitude takes, whether it is propositional or not. There is also an outside chance that pleasure should be understood as a behavioural disposition, which arguably would make it an easier thing to study scientifically.[4]

We are faced with a number of related phenomena: good mood, enjoyment, the feeling of well-being, pleasant sensations, pleasant thoughts, satisfaction. Ideally, we are looking for something that all these things have in common. There are differences between them, of course, because of the type of event referred to, but they do seem to have something in common as well that, arguably, is what an ambitious theory of pleasure should be concerned with.

The question I'm asking is not the question "what does the word 'pleasure' mean?", exactly[5]: the word 'pleasure', and its cognates, is used in a great variety of ways that relates semantically more or less closely. I'm not getting into a contest as to find the best fitting paraphrase of pleasure statements, an exercise that strike me as as futile as it is beside the point. 'Pleasure' is often used to refer to the *cause* of pleasant experiences, and there are a number of other "elliptic"

[1] Sidgwick (1981), Alston (1967), Brandt (1967, 1998), Frankena (1973), Feldman (1997a), Heathwood (2006, 2007). Gosling (1969) points out that the sensation view was a product of British Empiricism, and should not be viewed as the historical default view.
[2] Feldman (1997a), Heathwood (2007).
[3] Anscombe, (1967) see Katz (2006) and Crisp (2006).
[4] Gilbert Ryle (1969, 2000).
[5] See Perry's "the Concept of Pleasure" (1967), an exercise in ordinary language philosophy that spends a tremendous amount of effort listing the alternatives.

uses, such as when we say "pleased to meet you", which might truthfully be said while experiencing no feelings at all. Nor is the question under consideration "What is *happiness*?" Whereas I happen to believe that pleasure is the critical part of happiness, forms the core of that notion and is what is important about it, that term is imbued with too much meaning, too many preconceptions about the good life, to make a non-circular argument for hedonism possible. 'Pleasure', on the other hand, seems to be relatively free from such morally committing dimensions.

Pleasure is not only an everyday concept but one with use in scientific psychology as well. A satisfactory theory of pleasure, I propose, is one that fits not only with everyday uses of the term, but also with the best available scientific understanding of the domain. Ideally, such a theory would not only fit with such use, but make sense of it. We are at least partly interested in revising our everyday concepts to *improve* on them.[6] If there is a congruent class of scientific phenomena with which some philosophical theory of pleasure fits, that is a further reason to accept that theory. If we treat 'pleasure' as tracking not only an everyday concept, but a natural, psychological *kind*, the theory of pleasure should be done in conjunction with the affective sciences. It is in such a joint project we are most likely to cut nature at its joints.[7]

When we say that pleasure is important, we imply that it is not only the *essence* of pleasure that is of interest. That is, of course, of great philosophical and scientific interest, but we are also interested in what pleasure *does*, in its function and place in our psychology. The centrality of pleasure concern not only its essential, intrinsic features, but its typical causes and effects, the processes in which it takes part. All this influence how pleasure relates to motivation and action and sociality and to the rest of our psychological make-up. While not strictly essential, this project is every bit as important. For one thing, contingent yet persistent psychological connections can appear to be

[6] I take for granted that we are thus interested in getting our psychological language to chime with how our psychology works. This theme will recur in the next part concerning value.
[7] See Katz (2006), Berridge (2003, 2004), Kringelbach (2009), Schroeder (2004).

essential. If we want to get to the bottom of what pleasure is, we need to be able to distinguish such contingencies from essential features.

An account of pleasure needs to satisfy at least three conditions: it must give a plausible psychological picture that accounts for the apparent centrality of pleasure in matters like motivation and evaluation. It must be *phenomenologically* accurate: when it comes to subjective experiences, it is methodologically justifiable to ask about any proposed analysis of pleasure whether it actually fits with what we have in mind when we think of pleasure, to test whether we have caught the right notion or not. Finally, it must make it plausible that pleasure is good, i.e. it must fit with some plausible account of value.

This chapter starts with an outline of the main theories of pleasure and points out the challenges facing them. It ends with a suggestion of how those challenges can be met in a theory that incorporates the benefits of those theories, while avoiding the pitfalls.

1.1.2 Two Standard Views on Pleasure

It has become standard practice to distinguish between two main types of theories about pleasure. The first is the *Distinctive Feeling View* (DFV), according to which pleasures are experiences distinguished by a particular "hedonic tone" which they have and other experiences lack. The other is the *Desire Oriented,* or *Attitudinal, View*, according to which pleasures are experiences distinguished by some attitude that the agent has toward them.[8] This distinction is often treated as co-extensive with the more general distinction between *internalist* and *externalist* views on pleasure:[9] if pleasure is a sort of feeling, what makes an experience a pleasure is *internal* to that

[8] This distinction is in Feldman (1997a), the distinction is also made by, among others Gosling (1969), and Crisp (2006).
[9] This distinction is in Sumner (1996: see Crisp (2006)).

experience, and if what makes it a pleasure is a desire that one has towards it, that seems to be an *external* fact. The two distinctions are not necessarily equivalent, however: there are internalist versions of the desire-oriented view (but, to my knowledge, no externalist version of the distinctive feeling view).

In very short summary, the desire-oriented view was developed as a reaction to a fundamental problem for the DFV, namely the reported lack of such a distinctive hedonic feeling. The class of experiences grouped as "pleasures" is phenomenologically heterogeneous. What holds the class together and makes it interesting is something else. Sidgwick (1981), famously, argued that what we find in common between pleasures is not how they feel, but some attitude that we take up against them.

It is possible that there are *two types* of pleasures, in which case there really is such a thing as a distinctive feeling of pleasure, but that the term "pleasure" also denotes a distinct phenomenon, such as described by the desire view, and that the two only significantly overlap. Possibly, experiences having this feel were often desired, and thus the term came to cover all cases of desired experiences. There might also be other semantic connections between the two types.

1.2 The feeling of pleasure

> [Pain and pleasure] like other simple ideas cannot be described, nor their names defined; the way of knowing them is, as of the simple ideas of the senses, only by experience.
> Locke (1975, p 141)

Pleasures form a class of psychological events or states that are presumably not grouped together by accident. It has been proposed that what they have in common is how they *feel*. Locke goes on to say that pleasure and pain are not only simple ideas but "very considerable" ones.[10] Bentham, revealing similar sentiments, calls them "interesting perceptions".[11] Since much of the most influential writing on pleasure was performed during the heyday of British empiricism[12], this view has often been equated with the view that pleasure is a species of *sensation*.[13] A great deal of the criticism of the feeling view has therefore been based on the many ways in which pleasures are different from sensations.[14]

What is distinctive about experiences is that they are essentially conscious; that, in Nagel's terms, there is something it is *like* for someone to have them.[15] Experiences, in yet other terms, have a *phenomenal character*. This is not true of all mental states. Not all mental states are distinguished by how they feel, if indeed they feel like anything at all. What makes the belief that it rains different from the belief that it doesn't is a arguably not how the belief feels, but the *content* of those states, revealed by the inferences you tend to make. Note that

[10] He continues that words are not important, pleasure and pain might as well be called "delight" and, rather endearingly "trouble".
[11] Bentham (1960) .
[12] Bentham,(1960), Mill (1993). See Gosling (1969), Katz (2006).
[13] Locke, however, thought they derived from both perception and reflection. (Locke, 1975)
[14] Gosling, (1969), Feldman (1997a and b), Goldstein (1989) Alston (1967).
[15] Nagel (1974), Jackson (1982), Chalmers (1996).

this doesn't preclude that we can be phenomenally conscious of our beliefs: it only means that this is not what differentiates beliefs from each other. In contrast, what makes the experience of red different from the experience of green is how they "feel" in this sense.[16] Is there something it is like to experience pleasure? Is there some quality, "hedonic tone", that makes an experience one of pleasure, and thus makes pleasures into a cogent class?

1.2.1 The Phenomenological Component

What, if anything, can be said about the "essence" of pleasure, if it is a type of experience? As Locke pointed out, simple ideas (*qualia*, as they are now called) are basic and unanalysable.[17] But that does not mean that they cannot be intelligibly described. Locke himself described them as "very considerable". Arguably, they can be picked out via a description, by comparison or analogy, even if that description does not capture their "essence". While it does seem impossible to describe what it is like to have an experience to someone who has not felt it, nor something "like" it, we can remind people capable of the sort of experiences we are talking about of the right sort of idea, by an appeal to their typical causes or to situations in which they tend to occur. We can also circumscribe it by examples: the hedonic quality is that which all affective experiences, such as positive feelings, moods, sensations, have in common.

If pleasure is a type of experience, we can say something about what *kind* it is, especially with regard to the *generality* of that type. If we take the experience of colour as our preferred analogy: Is pleasure like some particular colour, or even a nuance of a colour? Or is it a more fundamental category, perhaps even as wide as the category of colour as such? However different experiences of colour are - some are even experienced as opposites[18] - there is something they have in common as to what type of experience they are, they are all in the same "mode", so to speak. Pleasures might be said to occupy a section of a dimension

[16] Let's for now pass over the question whether perceptual experiences are a form of belief.
[17] Moore (1993) called it a definite thing and *absolutely indefinable*, see Alston (1967).
[18] Plato in *Philebus* (1982).

or scale of some sort, on which experiences may then vary.[19] It seems clear that if pleasure has a particular, unanalysable, simple feel, it need not be simple in the sense that it is an on or off matter.

1.2.2 Problems For the Feeling View

A substantial part of the critique of the feeling view is that pleasure differs from ordinary sensations.[20] Gosling notes that standard examples of sensations are identified either by 1. their typical occasion or cause (i.e. a sensation can be identified as the feeling you get when you find yourself in a certain situation or encountering a certain sort of object), or by 2. what the subject feels like doing (i.e., a sensation can be identified by how we react to having it), or by 3. some analogous description, what the sensation is similar to. Pleasure, he argues, does not fit this schema. There are no standard occasions or sources for pleasure. People vary indefinitely in what they take pleasure in and "a person may be eccentric without limit in the sources of his pleasures". Pleasure is not associated with any standard behavioural response either, as this varies indefinitely between people and contexts as well.[21] And finally, pleasures cannot be understood in analogy to anything else. This, of course, does not prove that pleasure is not a distinct feeling, but it shows that it doesn't work entirely as we would expect it to if it were a sensation.[22]

In contrast to sensations, pleasures are *second order* experiences. That is, they are not direct perceptions, but *reactions* to some experience. While this undermines the *sensation* view on pleasure, it provides it with another role: not all phenomenal experiences are "first order". This justifies locating it among the *emotions* rather than the sensations (more on this below). Gosling argues that

[19] Kagan (1992). See Crisp (2006) Katz (2005).
[20] See Gosling (1969), Alston (1967), Feldman (1997a).
[21] Whether to go for it or stay put, for instance. Similar point made by Persson (2005). Of course, this holds for most sensations to. You do not need to score on each of these points in order to qualify as a sensation.
[22] Momeyer (1975) equally pointed out that sensations and pleasure work with a different logic.

feelings of pleasure are properly conceived as emotional responses to some experience. Pleasures are not mainly feelings *of* pleasure, but of something else: pleasure has an experience as its *object*, to which the pleasure somehow attaches. Gosling reminds us that pleasure often makes us attend, not to itself, but to the thing we are doing or experiencing. Whereas the intensity of a sensation makes it salient, intense pleasures tend to make the object salient, not itself. This, however, doesn't undermine the feeling view. It only shows that the pleasure is not the object of that state.[23]

The objects of sensation are often external to the agent, whereas the object of pleasure is often a sensation, or other subjective state. Denial of the status of sensation or perception to pleasures seems to be based on the fact that the information it provides is not as objectively valid as it normally is for sensation/perceptions. Pleasures are not subject to tests for reliability in the way that our sensations are. Pleasures' variability rule them out as sensations/perceptions proper, but it would be a strange view indeed that took this to undermine their status as *experiences*, i.e. as essentially subjective events.

A further argument against the sensation view of pleasure is that pleasure lack a localisation. This also undermines the analogy between pleasure and pain since the latter often *is* localised.[24] This is because pain actually *is* a sensation, at least one part of pain is.[25] There is a suffering element to pain that is as non-localised as pleasure is, but there exists no distinct analogous sensory dimension of pleasure. In so far as we speak of sensory pleasures, it refers to their source, not their location.[26]

---

[23] Persson (2005) argues that an experience is never *only* painful, or pleasant, but always something else as well. Duncker (1941) argues that pleasure is incomplete.
[24] Momeyer (1975), Alston (1967).
[25] See Aydede (2000) , Melzack and Wall (1965).
[26] This is not beyond doubt. Some people (among them, at least one of my supervisors) seem to experience, not only the cause of, or object of, bodily pleasures as localised, but the pleasure itself. It is hard to say whether this "disagreement" refer to fundamentally different experiences, of just different ways of describing the same experience.

Another argument claims that pleasure cannot be a sensation because every sensation can be either pleasant or unpleasant.[27] Ryle notes that any sensations may monopolize consciousness, if intense enough, but the intensity of pleasure only serve to increase the consciousness of the thing we take pleasure *in*. Alston similarly claims that we cannot have the pleasures of x, without consciousness of x. Pleasure is not "detachable" from the experience it accompanies. Again, these points merely demonstrate that pleasure is not a sensation, they don't prove that it is not an experience.

Moore argued that since we can be conscious *of* pleasure, pleasure must be distinct from our consciousness of it. If this argument is supposed to undermine the feeling view, it is easily met: pleasure is not merely an independent *object of* consciousness: it *is* a state of consciousness. We can certainly experience pleasure without having a second order awareness *that* we have it, but that does not undermine the feeling view in any way: we can be conscious of x, without being aware *that* x occurs. Moore uses this argument to undermine, not the plausibility of this theory of pleasure, but of *hedonism* by noting that pleasure without the consciousness of pleasure seems to be of comparatively little value. Seeing how "consciousness of pleasure" can be understood in two ways, this argument is weakened.[28]

Some writers argue that there exists no dedicated organ or faculty for pleasure, as the ones we find for the senses.[29] Alston argues that this means that there is no "external support", no modality or organ or stimuli dedicated to pleasure, "nor can anything much better be found on the response side". Despite there being some significant overlaps in the kind of things people get pleasure out of, it is not enough for an organ to be selectively dedicated to registering it, and for pleasure to be thought of as a reliable indicator. In recent years, however, the

[27] Ryle (1969, 2000) see Alston (1967).
[28] Moore (1993).
[29] William James, for one (1950). The lack of fair treatment of pleasure in James' hugely influential work is probably partly responsible for the decline of the hedonism in the 20th century. This was when the ties to psychology were severed and, as I'll argue in part 2, hedonism is dependent on such a tie.

rise of movements like positive psychology, happiness research and affective neuroscience has led to the discovery that there is such a faculty, roughly localised in the orbitofrontal cortex of the brain.[30] There is no dedicated *sense*-organ for pleasure, however, but the existence of this region of the brain, that can be selectively targeted, does provide the "external support" that Alston reported missing.

The sensation theory, Alston recognises, is merely a variant of a more general sort of view that takes pleasure to be one of the "ultimate immediate qualities of consciousness/experience. To be a quality of consciousness is to constitute one of the ways in which one state of consciousness differs from another with respect to its own intrinsic nature. It is noteworthy that Alston finds this theory implausible too, but on purely phenomenological grounds.

While pleasures are not exactly like sensations, there is a case to be made that pleasures can be identified in the manner proposed by Gosling. While there is considerable variability in the kind of objects and situations we find ourselves enjoying, there certainly are some standard examples of pleasant activities, and there might be similarities on *some* level. One such suggestion is that pleasure results from getting what we want, and while *what* we want might vary indefinitely, they are all occasions of getting what we want. While sensations are often held to be objective in the sense that they provide publicly available information about an object, that a thing is wanted by me is clearly relevant information, well worth a particular mode of experience. As to the response side, that is in all probability dependent on what kind of need or attitude has thus been satisfied. The last point on Goslings lists of complaints was the lack of analogy. But pleasure is unlike anything else because it is too generic a category for it to be understood via analogy: it is a *sui generis* kind of experience. In the same way colour, as opposed to some particular colour, has a distinct character, unlike anything else. Pleasure is what all positive emotions have in

[30] See Kahneman et al (1997), Nettle (2005) Berridge (2003, 2004), Panksepp (1998) and Kringelbach (2001, 2009), Bressan and Crippa (2005) Crisp (2006) makes this same argument.

common, and positive emotions can be understood with reference to each other, but what makes the category as such distinct cannot be understood other than by knowing it directly.

The heterogeneity argument

Now, let's turn to the main argument against the feeling view. This argument is associated with Henry Sidgwick in *the Methods of Ethics.*[31] Quite simply, it is the claim that the experiences classed as pleasures have nothing phenomenologically, or intrinsically, in common. They are *heterogeneous.*[32] The pleasure of listening to a Mozart opera, say, feels nothing like the pleasure of slipping into a hot bath on a cold day. This is not just because of the different modes of sensation involved: the pleasure of listening to a Mozart opera is arguably distinct from the pleasure of listening to John Coltrane as well. Whatever is distinct about pleasures, then, it is not how they feel. If there is anything to be salvaged from the talk about the "feeling of pleasure", it is that they are distinguished by how we feel *about* these experiences. That locution is not supposed to express a feeling, exactly, but rather a *sentiment*, a favourable *attitude* towards the object or state of affairs enjoyed.

The heterogeneity argument draws most of its strengths from the appeal to the different activities and objects one may take pleasure in, and the wildly various experiences that these activities and objects afford. "Pleasure", normally, refers to entire experiences so that at least one of the differences between the pleasure of listening to Mozart and listening to Coltrane is that their music *sounds* different. It is not merely that the pleasure of listening to Mozart is one that occurs simultaneously with the experience of listening to Mozart:[33] the pleasure and the experience are more closely knitted than that. The pleasure of listening

[31] Sidgwick (1981).
[32] The argument has been assigned to him by Brandt (1967) Feldman (1997), Sobel (1999) among others. But there is an ambiguity in the central statement of his view on pleasure as "desirable consciousness".
[33] As Momeyer (1975) points out, the pleasure of playing tennis *implies* the experience of playing tennis. Alston (1967) also addresses this as the *binding* problem.

is not distinct from the experience of listening. That is why it cannot be understood as a separate sensation.

This still leaves the question what this attitude actually *is* unanswered. What kind of attitude is it? Is it something that is felt? In that case, this is still a version of the feeling view. Unfelt attitudes would not transform an experience into a pleasure.[34] The heterogeneity argument, if successful, need to say something stronger than just that pleasure is a heterogeneous set of experiences: it needs to say that pleasures have nothing phenomenologically in common. While many theorists have accepted this, it is far from clear that Sidgwick did, as we shall see in the next section.

Is pleasure always felt?

A quite different challenge to the feeling view is the claim that pleasures are not necessarily conscious.[35] There are two points here: even if pleasure always has an effect on the quality of our experience, this need not be *noticed* by the agent. Arguably, our conscious experience has a large number of features that we do not normally *attend* to, and yet they are there to make up the complete character of our experience. As pleasures are often experiences of other objects, and those objects make up the *focus* of those experiences, pleasantness often goes unnoticed. In fact, as critics of the sensation view noticed, increased pleasantness has a tendency to increase the attention paid to the object of an experience, rather than to the experience itself.[36] This point is quite compatible with pleasure being essentially a conscious quality: the fact that pleasure is conscious does not imply that, when we experience pleasure, we are always conscious *of* that fact.

Second, whether or not you conceive of pleasures as essentially conscious, they depend on the existence of some functional, neurological state of the organism.

[34] Sobel (1999) thinks there is no middle position, as suggested by Katz (1986) and Kagan (1992).
[35] See Berridge (2003, 2004) Persson (2005).
[36] The argument is that intense sensations crowd out consciousness of anything distinct from it.

Pleasure can be, and has been, operationally defined; notably as *unconditioned reward*, i.e. that for which the organism is willing to work.[37] The point is that the same process can occur below the "threshold" of consciousness.[38] It should be pointed out that it is not clear what this metaphor of a threshold actually entails, but if this is a possibility, would such a state *count* as pleasure? The matter seems to be dependent on what we are interested in.[39] It can be argued that since we *identify* this functional/neurological state by how its full-fledged version feels, that feeling is at least *epistemologically* prior, while the process might be *ontologically* prior. In contemporary affective science, both the operational, functional view, and the experiential view seem to have a strong standing, and they are not mutually exclusive.[40]

[37] See Schroeder (2004) Berridge (2003) and Kringelbach (2005). Of course, finding such a basis was important during the *behaviourist* era (see Ryle 1969).
[38] This threshold view of consciousness is quite common. See Ledoux (1996).
[39] See Chalmers (1996) on the "hard" and "easy" problems of consciousness.
[40] See Kahneman, Diener and Schwarz, (1999), Kringelbach (2005, 2009), Berridge (2003). Kahneman (1999) points out that moving away from experienced utility towards behaviourally oriented research, as happened in economics.during the 20th century, is problematic. While it is easier to measure, this move misses the point. Experienced utility, in fact, is both measurable and empirically distinct from decision utility. Momeyer (1975) understands pleasures as dispositional states: states that *would* be experienced as pleasure if attended to.

1.3 The desire oriented view

How, then, should we distinguish the class of pleasures, if we give up on the distinctive feeling view? According to the theory that usually is offered as an alternative, pleasures are experiences for which we have some favourable *attitude*.[41] It is certainly a fact that we are normally drawn to pleasure, and repelled by pain. We often use pleasure to *explain* desire: we come to desire things we find pleasant or expect to get pleasure from.[42] We also use desire to explain pleasure: we are pleased by the outcome of an election or by the taste of ice cream because we desired that outcome or that taste. These sorts of statements make sense of particular pleasures and desires, and the kind of explanations they offer seem to be part of the folk-psychological toolbox.

In the absence of a distinctive feeling of pleasure we can turn to this fact, and treat it as the distinguishing feature of this otherwise motley class of experiences. In "The methods of ethics" Henry Sidgwick defends a version of this theory. He suggests that pleasures be conceived as experiences for which we have an intrinsic *desire* at the time we experience them.[43] This formulation already includes four important qualifications of the attitudinal theory. First, the pro-attitude in question is *desire*. We shall return below to what this involves. Second, the object of the relevant desire is an *experience* that the subject has. Third, the relevant desires are *intrinsic* ones: we often desire experiences for instrumental reasons, but those experiences do not thereby count as pleasures. Fourth, the desire must be simultaneous with the experience. This is to insure the account from cases of disappointment, where an intrinsically desired experience turns out to be less than hoped for. Further

[41] See Fred Feldman (1997a), who holds this to be the new standard view, citing Brandt (1967), Alston (1967) and Frankena (1963).
[42] As William Alston puts it (1967) "It seems clear to most people that pleasure and enjoyment are pre-eminent among the things worth having and that when someone gets pleasure out of something, he develops a desire for it."
[43] Sidgwick's statement that pleasure is "at least implicitly perceived as desirable in it self" is rather more open for interpretation, but he is most often read in this way.

qualifying, or clarifying, this view, William Alston suggested that the relevant desire is a preference for an experience over its non-occurrence *on the basis of that experience's felt quality.*[44] This, of course, follows if all intrinsic features of experiences are qualitative in that sense. Brandt, continuing the same tradition, suggests that an experience is pleasant if it makes the person experiencing it want its continuation (for its own sake).[45]

This view is not the claim that pleasures are as a matter of contingent *fact* picked out by intrinsic desires, i.e. that this is how to identify them. As Alston points out, the fact that pleasure is desirable does not seem to be a mere contingent matter. The feeling view, he continues, can throw no light on this fact. Nor does the fact that the enjoyableness of an activity is a reason for doing it seem contingent.[46] If you have a desire-oriented view of the good, the desire-theory of pleasure explains why pleasure seems to be notoriously good. Making the connection between pleasure and favourable attitudes an essential one makes hedonism more attractive as a theory of well-being, and as a theory of the good.[47] The theory is also to be kept apart from the claim that pleasure is the *only* thing we desire.[48] Any *experience* we desire in the relevant way will thereby count as a pleasure, but this does not bar us from desiring other things for themselves, without those things thereby becoming pleasures. Whether or not the desire-view of pleasure in conjunction with a desire-theory of the good supports hedonism or merely the value of pleasure *among other things* ultimately depends on how we construe the relation between pleasure and desire, and between value and desire.[49]

---

[44] Alston (1967)
[45] Brandt (1998). Since Alston's view on preference is dispositional, he arguably intended something similar.
[46] Alston (1967, p345).
[47] Gosling (1969), believes that hedonism depends on a connection from pleasure to rational, free action. A desire-version of hedonism, then, as opposed to the objective list version on the DFV, see Kagan (1992).
[48] Indeed, Sidgwick is known for rejecting psychological hedonism. Alston concurs (1967).
[49] Heathwood (2006) argue that the most plausible version of the desire-view is identical to the most plausible version of hedonism.

1.3.1 The Desire-component

Before assessing the desire view and its varieties, we need to make some general remarks on what desire is. Without proposing to settle the matter, or giving anything like a complete survey of the literature on the subject (which is vast), there are some preliminary remarks we can make.

'Desire' is normally used as a general term for pro-attitudes.[50] While admiration is quite obviously a distinct mental act from fondness or love, they are all favourable attitudes, and 'desire' is often used as a catch-all term for the species. As with the term 'wanting', with which it is often conflated, 'desire' is often used to explain free, rational actions. Why was that action performed? Because the agent wanted to do it, or desired the outcome. While 'want' and 'desire' can be used interchangeably, they can also be contrasted, and someone may intelligibly ask whether I want what I desire. Indeed, one may intelligibly ask whether I desire what I desire, suggesting that 'desire' stands for a cluster of related phenomena, all being cases of favouring, but that they can come into conflict within an agent. What we mean with this contrastive use of the same term is normally conversationally implied.

Two views of desire

Desire can be understood in at least two different ways: as a dispositional state, or as an experience.[51] When desires are used to explain action, they are normally conceived of as *dispositions* to act. If you *really* desire something, you will tend to bring it about, if it is not already a fact, or to preserve it, if it is. Failure to comply will undermine our confidence in assigning you the desire. If desire is a disposition, it is a certain sort of disposition, a tendency to perform the action *willingly*, which makes it distinct from reflexes or forced behaviours.[52]

---

[50] Heathwood (2006, 2007) takes it to be a "primitive" and uses it as "the paradigmatic "pro-attitude"."

[51] See Sidgwick (1892).

[52] We do seem to say that we have reluctant desires, urges, that exist somewhere between rationally willed action and mere reflexes, and there is arguably no sharp line dividing the two.

William Alston, in the entry on 'pleasure' repeatedly referred to, talks about *preference* rather than desire, and notes that to have a preference is not necessarily to have it before one's consciousness, but rather to say something dispositional. We have access to our preferences in the same way that we have access to our beliefs, intentions and attitudes, "as well as to feelings and sensory qualities". Note, however, what Alston doesn't: that these are distinct forms of "access". Ned Block calls them "*phenomenal*" and "*access"* consciousness, and to have *access* to a mental event is not the same thing as to have it in one's phenomenal consciousness.[53] What Alston's account guarantees, however, is that the epistemological status of pleasure – if we have it, we know that we do – is compatible with the view that pleasures are not necessarily felt. Or, at least, it grants the same sort of epistemic access to our pleasure as it does to our beliefs and attitudes.

Heathwood points out that many philosophers adhere to the principle that we cannot desire/want what we already have, which undermines the desire view that *requires* the desire to be simultaneous with the experience desired[54]. Heathwood denies this principle. Clearly there is *some* pro-attitude we can bear towards things that we have, and this pro-attitude is included in what he intends with "desire".[55]

Problems for the dispositional view of desire

There are problems for the dispositional view: might I not favour things that I have no disposition to bring about or preserve? There are things that I favour that I can do nothing about. Perhaps desires should be understood as dispositions to do something to bring it about *if* it was possible, but this makes little sense when applied to desires for things that are clearly impossible, say, or that has happened in the past.[56] Of course, we *can* formulate such conditions,

[53] Block (1995), for instance.
[54] Heathwood (2006) Sumner, for instance, argues that desire is "essentially prospective" (1996).
[55] Perry (1967) agrees: there need be no tendency to linger, nor a pre-existing desire in order for you to enjoy something.
[56] For extended treatment of this argument, see Strawson (1994) and Schroeder (2004).

but they don't seem to be what we have in mind when we think about our desires. We seem to have favourable attitudes about things that no one can do anything about. Of course, very often, we would have brought about a desired state if we could have, but then it seems that the desire is what explains that counterfactual, rather than being identical to it. Galen Strawson invented a hypothetical type of being he called "Weather watchers" who, deprived of any type of capacity for action, still could have a desire for how the weather turns out, and I see no reason to rule it out.

Problems for the experience view of desire

We sometimes speak as if we *feel* desire. Could the experience of desire be the feeling *of being in the relevant dispositional state?* Perhaps it is the conscious *representation* of the desire, and thus distinct from it. That would explain why the same term is used for both, and it would be a matter of decision rather than discovery whether we should treat the disposition without the experience as a desire or vice versa. This would also provide us with the tools to deal with desires that are *not* coupled with actual dispositions: one of our remarkable mental skills is the ability to represent what is not there. Again, the availability of two distinct phenomena make contrastive uses possible: when asked whether I *really* desire something, I might be questioned on basis of my reluctance to actually do something to bring it about, or I might be questioned on whether I really *feel* like doing what I'm obviously disposed to do.

This possibility, however, seems to presuppose that there is a homogeneous type of experience that represents the dispositional state. But how a disposition feels depends on what it is a disposition for: being ready to dive into the cold water on a hot day feels quite different from getting ready for bed when tired, or just disposed to keep on doing whatever it is that one is doing. There is a heterogeneity argument for desires too, obviously, but it is one we can get around. What is in common for them is that they are all states of *readiness*: their similarity is on a higher level of generality than their particular physical manifestation.

More importantly for our purposes, however: some desires are pleasant, others are painful, and this is not just the difference between the experience of satisfied desires and the experience of a dissatisfied, prospective one. While we might rarely, if ever, experience (known to be) satisfied intrinsic desires as unpleasant, unsatisfied ones can be either. But if the desire view of pleasure is true, how could we make sense of pleasant and unpleasant desires? Presumably, an experience for which we have an intrinsic unpleasant desire would not thereby become pleasant. The desire theorists might propose that an unpleasant desire is one that we do not desire to have. While that sounds about right, its conceivability depends on how *that* desire in turn is understood, i.e. as a disposition or as a feeling, and the problem is merely deferred, not solved.

If there are two different senses of desire, which one is relevant to pleasure?[57] If desire is a disposition, the theory runs into certain problems. If it is an experience we seem to run aground on the heterogeneity problem again.

Type of object

Let's turn to another matter of contention for the desire theory. What kind of an *object* does desire take? An influential suggestion is that desires, like beliefs, are *propositional attitudes*. Whereas we sometime speak as if we desire objects, a new car, say, or true love, those expressions are elliptical for propositional objects. What we desire is *that* we get a new car, or *that* we be seen driving around in it; *that* we attain true love or something of that nature.[58] This interpretation is in keeping with the dispositional view. You cannot bring about or preserve an object with out bringing it about *that it obtains*.

Now, if this is true, it seems that experiences *can't* be the object of desires, in the sense required by the desire-oriented view of pleasure. Whatever experiences

[57] Gosling points out that Mill's view about the conceptual/metaphysical connection between pleasure and desire/wanting mistakingly supposes that "want" is just one, single thing.
[58] Feldman (1997a), Lemos (1994) Parfit (1984).

are, they are not propositional in form. Some philosophers have denied that desires are propositional attitudes on precisely these grounds: we occasionally favour an object without thereby favouring *that* it exists[59], and if the term 'desire' does not cover such pro-attitudes, then we need to turn to the more general notion to cover the cases we are interested in. This might be a good idea anyway: Favoured experiences - and already attained states of affairs - are perhaps more fittingly described as *liked* or *enjoyed* than as desired.

Experiences are not states of affairs, but concrete objects/events.[60] Now, we might point out that desires have so called *mediate objects*, i.e. a representation of their object, and that this mediate object is always propositional in form. You cannot imagine an object without some predicate, even if you claim to desire a concrete object; that state of affairs is what you "truly" desire. Even if that is plausible for most cases, there is still one type of object that *doesn't* require any mediate representation, namely experiences. Since desires and experiences are both mental events, they would seem to need no representation to mediate between them. For experiences it seems quite clear that favouring *it* is distinct from favouring *that* one has it, even if this distinction could be denied for any other object of desire.[61]

The temporal placement of the desire

As mentioned, the relevant desire needs to be simultaneous with the experience. This is to get away from hedonic disappointments and to allow for pleasant surprises. Chris Heathwood describes a case in point[62]: I might have a strong intrinsic desire for some taste experience I had as a child, like the taste of *fruit loops.* But when I get hold of them, it turns out that they are far too sweet for my refined tastes. A previous desire does not insure that the taste will be pleasant, what is important is that we have a desire at the point that we have the taste. It is also important that the desire be somehow *connected* to the taste. I

[59] Anscombe (1967) Katz (1986, 2006).
[60] See Rønnow-Rasmussen (2002).
[61] Katz (1986, 2006), Anscombe (1967).
[62] Heathwood (2006).

might be experiencing a desired taste-experience, but not realise that this is the taste that I desire intrinsically. That, arguably, would not be enough to make that taste pleasant. Heathwood therefore adds that we must be *aware* of the sensation in question for a concurrent desire with it to be a pleasure. I agree, but propose that the awareness implied need not be the awareness *that* the desired experience is happening. It must be awareness *de re*, and not de dicto: the desire must be about the sensation *itself*, not merely the belief that one has it. That is: Awareness *that* I have the experience is neither necessary nor sufficient. We must be directly acquainted with the experience in question.[63] But In Heathwood's formulation (see below), the desire does not even have the sensation as its object, but some proposition in which the sensation is represented. This would lead the desire-view into trouble with ensuring a sufficiently tight connection between the desire and the pleasure experience.

1.3.2 Problems For the Desire View

The desire view faces a number of difficulties, some of which are theoretical and some are more directly intuitively based. If we keep to the original formulation of the desire-view, namely that pleasures are experiences for which we have intrinsic desires, there are two clear deal-breakers: intrinsically desired experiences that we would not call pleasures, and pleasures for which we have no intrinsic desire. Plausible example of such events would be clear Socratic evidence that the definition we are considering is a faulty one.

Other reasons for intrinsic desire

Are all intrinsically desired experiences pleasures? What about experiences that we just find *interesting*, and intrinsically so? Might we not desire, intrinsically, to have them, without that making them instances of pleasure? Heathwood argues that such an interest actually *would* make them pleasant, but his argument is based on the plausibility of the theory he is proposing, and thus offers no *independent* reason for the claim. In particular, it is dependent on the

[63] In Alston's formulation, the desire must be for the experience "For how it feels". This should be understood de re.

feeling view of pleasure being false. Heathwood argues that giving *reasons* for a desire is evidence of the externality of that desire, but that seems just false.[64] That something is interesting need not be an extrinsic consideration: we can give *internal* reasons for interest. On the desire view, we cannot get around this problem by adding the condition that the experience be desired because it feels *pleasant*. And that is a bit odd, because that just seems to be the best reason to desire an experience because of how it feels.[65]

Remember that it is essential that the desire appealed to is *intrinsic*, i.e. that an experience is desired because of intrinsic features of its object. But if the object is not the experience, but some state of affairs in which the experience is mentioned, why is it still important that the state of affairs is desired because of features intrinsic to that experience, i.e. how it feels, rather than to the states of affairs in which it is involved?[66] On versions of the desire view that defend a propositional conception of desire, it would seem that the experience is itself not the *object* of the desire but only included, or even merely *mentioned,* in that object. And yet, it would not seem to be sufficient that an experience is included in an intrinsically desired state of affairs for it to become a pleasure. The propositional desire view owes us an explanation of why this is so.

Explaining desire

A further theoretical difficulty for the desire theory is the matter of *explanation*. While we often say that we desire an experience for the particular taste or sound it presents, at least occasionally we desire an experience because it is *pleasant*. This suggests that pleasure can be *prior* to desire. But if what *makes* it a pleasure

[64] Heathwood (2007) believes that to desire something for its intrinsic qualities is distinct from desiring it intrinsically. Feldman, (1997a) thinks it's possible to desire intrinsically to be feeling some sensations without that sensation being a sensation of pleasure. The pleasure is a *propositional* thing, not a sensation.

[65] More obviously, perhaps, some undesired experiences are not pain, and this is probably correct, since pain is not the opposite of pleasure. Displeasure and unpleasantness better fits this description. See Rachels (2004).

[66] Heathwood (2006) offers his theory that a sensation, occurring at time t, is a sensory pleasure at t iff the subject of S desires, intrinsically and *de re*, at t, of S, that it be occurring at t. The sensation is not the object of the desire.

is a desire this claim seems circular. The claim is only circular if the desire explained is the *same* desire as the one that makes it a pleasure. I might desire an experience because there is a quite *distinct* attitude that makes it a pleasure. There is in fact nothing strange, or even unusual about liking things *because* we like them. We merely need to keep in mind that there is more than one attitude at play when we explain desires in terms of pleasures. In addition, if the desire theory is true, the circularity never arises. Since what makes something a pleasure is extrinsic to the experience, desiring it because it is pleasant is never to desire it for intrinsic reasons. Whether the desire view is compatible with the explanation of desire in terms of pleasure is thus dependent on the plausibility of adding a further attitude to the mix. I will leave that to simmer for a bit, and we'll return to this suggestion further down.

Non-Intrinsicness

According to the desire view, what makes an experience a pleasure is something *extrinsic* to that experience. As pointed out, the desire view is guided by the conviction that what determines whether our experience is a pleasure or not is not how it feels. Since, arguably, everything intrinsic to an experience is a fact about how it feels, the desire view is contractually obliged, as it were, to deny that experiences are pleasures in virtue of their intrinsic features. But this brings us into trouble if we wish to say that pleasures are *intrinsically good*.[67] What makes pleasure good is arguably what makes it a *pleasure*, but, if the desire view is correct, that means that what makes it good is external to the good. While not all people think all pleasures are intrinsically good, surely most people agree that some pleasures are intrinsically good. It is surprising, Feldman notes, that so many hedonists have found the desire view compelling, seeing how it makes their position inconsistent.[68]

[67] Fred Feldman formulated this problem in an influential (1997a) paper, and offered a solution to which we'll return to later on.
[68] Feldman (1997a).

The argument hinges on a questionable premise, namely that pleasure has *intrinsic* value. In recent years, the idea that not all non-instrumental values are intrinsic values, but that there exist such a thing as extrinsic, non-instrumental value, has received sympathetic attention. Things might be valuable because of some relational property, like the property of being unique or significant, or the property of being created, owned or given by some particular person. This notion is often dubbed 'final' value.[69]

We certainly value things for their extrinsic properties, and for non-instrumental reasons. This does not mean that any of them have any value. Indeed, the fairly self-explanatory "isolation test" devised by Moore might be taken as a device to weed out precisely these sentimental or association-based goods.[70] But that is a substantial claim in need of independent support, and we shouldn't rule the possibility of such values out.

It doesn't much matter what we make of the examples proposed in that literature[71], since pleasure, if the desire view is true, offers the best possible argument for the existence of non-intrinsic final value. Pleasure certainly has value as an end, so if it cannot have intrinsic value, it must have final value. Let's just consider what this entails: Certain experiences are good, but *conditionally* so. They are not good in themselves, since a qualitatively identical experience that would *not* be the object (or "included" in the object) of the right kind of attitude would not be good, but might instead be neutral, or bad.

In general, if desire and/or response-dependency accounts of the good are correct, the availability of non-intrinsic, final value makes it possible for the adherents of such accounts to say that yet, it is the *objects* of those desires/responses that are good.[72] In defence of the notion of intrinsic value, we can argue that what is good in those cases is not the object, but the states of

[69] See Korsgaard (1983), Kagan (1998) Rabinowicz/Rønnow-Rasmussen (1999).
[70] Moore (1993).
[71] While the cases might be unconvincing. they are not based on a conceptual confusion.
[72] See the discussion of preferentialism in Rabinowicz/Österberg (1996).

affairs that include the object and the response.[73] The *reason* behind this move is the desire to include in the valuable objects everything that is *important* to their status as such, and since what determines the value of an object is not included in the object on this reading, it fails to do what we wish the theoretical notion to do.[74]

On a very similar note, a complaint against the desire-view is precisely that it does not include in the pleasure that which makes it a pleasure. If we are looking for something like the "intrinsic essence" of pleasure, it turns out that there is none. Of course, this is not unheard of, there is nothing strange in the notion of essential properties being external, relational (the essence of being a father, say, or a king), but it does seem to come at a price. If what makes an experience a pleasure is an attitude, why say that the experience is a pleasure? If we can, why not say that the pleasure *include* the attitude? It seems preferable that pleasure should include whatever makes it so.

Too demanding

Another objection targeting the narrow, propositional version of the desire theory is that it demands too much cognitive capacity. In general, agents incapable of anything as sophisticated as a propositional attitude with an intrinsically discerned object are yet capable of experiencing pleasure. Most animals and small children are obviously capable of pleasure, and yet one is hard pressed to conceive of them having propositional attitudes of this quite complex sort. Even for agents capable of entertaining such thoughts, it doesn't seem to correspond to what we are doing when experiencing pleasure. The two proponents of this view considered, Feldman and Heathwood, are generally very clear on this point: what ever these attitudes are, children and (most) animals are capable of having them. It is therefore disappointing that neither of

[73] This, indeed, is the defence Feldman uses in his (1997a) paper

[74] Of course, not everyone agrees about the desirability of thus including the critical properties in the valuable object, as the argument for extrinsic final value makes clear. The "argument" here depends on a commitment to a particular conception of intrinsic value.

them offers a theory of what desire actually *is*. This makes the claim difficult to assess.

Not plausible, if dispositional, not distinct, if it is a feeling

If the "desire" implied is a disposition, the analysis seems implausible, because even if behaviour, and thus dispositions, *are* used as indicators of emotion, in emotion research, as in everyday life, it is not fool proof.[75] Dispositions are highly unreliable indicators of whether some one is experiencing pleasure or not. How people behave when pleased seems highly individual. We do occasionally see pleasures revealed in people's behaviour, but it is not necessary to exhibit any particular behaviour, we need not even be disposed to prolong an experience we judge as pleasant.[76] If the desire is a form of *experience*, on the other hand, so that an experience is a pleasure if we experience some sort of attitude towards it, the view as expressed faces the same problem it was set to solve, and threatens to collapse into a form of the feeling view.

Pleasure and displeasure feel alike

The to my mind most decisive objection to the desire view is that it claims that pleasures and non-pleasures may feel the same. On that view, the difference between pleasure and pain *is not how they feel*, but what attitude we have toward them, and for what reasons. This means not only that what I experience as unpleasant, you may experience as pleasant, but that there is no intrinsic difference between those experiences. There is, of course, a lot of interpersonal overlap as to what kind of things we enjoy, which could account for a lot of the initial implausibility of this suggestion, but it does not seem to catch all of it. While it is true that people vary in what experiences they strive for and enjoy, what they get out of those experiences differs from what others get. The difference between listening to something and getting pleasure out of it and listening to it and being annoyed, for instance, is not necessarily a difference in

[75] See for instance Ledoux (1996) Sobel (1999).
[76] In addition, the desire to prolong an experience has a *future* object: that *this* experience *continues*.

how it *sounds*, but it quite clearly is a difference in how it *feels*. That's why some people tend to avoid skydiving, and others are drawn to it: the same physiological sensations are experienced as pleasant by some, and unpleasant by others. The difference in experience seems to explain our extrinsic attitudes and desires, rather than the other way around.

1.4 Pleasure as Representation

Regardless of whether we understand pleasure as a sensation or an emotion there is an alternative to the views considered, namely to treat it as a *representation.* The idea is that experiences have *content* that represents something, a state of the world or a state of the agent. Perhaps this is where the answer is: what pleasures have in common is what they *represent*. Let's start, however, with some notes about the content of emotions.

Some notes on cognitivism in emotion theory

According to the *cognitivist* tradition in the philosophy of emotion, emotions are a form of judgment. This idea has its offset in the observation that emotions/feelings are intentional, content bearing states.[77] Insofar as they are reactions to stimuli, they are not *mere* reactions: they say something about the stimuli. Fear, for instance, "says" that the stimulus is dangerous, and to be avoided. Emotions, according to this theory, are somewhat like beliefs: distinguished from each other by their propositional content. They are not a species of beliefs, however. Beliefs are dispositions, whereas judgments are more akin to *acts*. (Beliefs can, of course, be *manifested* in judgments). While intimately associated with "cognitive" judgments, emotions can also go *against* our judgments, as is the case in most phobias: i.e. the emotion judge as dangerous something that we know is not. This, the cognitivists reply, just means that two contrary judgments can be held at the same time. In addition, if an emotion can go *against* a judgment, it must itself be a judgment.

One influential suggestion[78] is that emotions are *evaluative* judgments: the stimulus is not merely categorised, but also evaluated as good or bad, and it is with this notion that (pure) cognitivist views struggle. It's hard to see what

[77] See Helm (2002), Katz (2006), Solomon (2003).
[78] Solomon (2003), Helm (2002).

the evaluative element of the emotion, which seems to be essential to it, would represent. It has been argued that the cognitivist theory cannot account for the *affectivity* of emotion.[79] Solomon, defending the cognitivist view, admits that no amount of information is sufficient to constitute an emotion: if that were the case, emotions could as well be beliefs. Emotions are, at least in part, experiences. Cognitivism merely claim that some experiences constitute judgments.[80]

1.4.1 The Matter of Representation

We will focus on a more specific part of the content of experiences, namely the issue of *representation.*[81] Representationalism, as distinct from cognitivism, allow for *non-propositional*, possibly *non-conceptual* content, which means that not all representational states are judgments *that* something is the case.[82] If pleasure shall be understood as a representation it might very well be of this kind. Pleasures, as we said, are often properly conceived as *reactions*. This opens up for the suggestion that they have propositional content: that they make some sort of claim about the stimuli. But this is not necessary. If they are indeed evaluative, they need not be understood as judgments *that* the stimulus is good, but may be *representations of the goodness of the stimulus*.

What does it mean that an experience represents something? Tye offers this short and snappy characterisation[83]:

> Experience represents various features by causally correlating with, or tracking, those features under certain optimal conditions.

---

[79] But see Solomon (2003).
[80] See the *appraisal theory of emotion* in chapter 2.5.
[81] Zajoncs (1980) point out that "preferences need no inferences". The appraisal theorists tended to disagree, but modern appraisal theories seems to invoke no explicit cognitions necessarily See Scherer and Ellsworth (2003).
[82] This distinction was brought to my attention by Marie Lundstedt.
[83] Tye (2005), See Also Chalmers (2004).

Does pleasure have representational content in this sense? Does it correlate with and/or track anything under certain "optimal" conditions? While this suggestion has received limited treatment as an account of pleasure, it has a relatively strong position as a theory of pain.[84]

1.4.2 The Illustrative Case of Pain

Representationalism about pain is more plausible than it is for pleasure, as the sensation model is more plausible for pain than for pleasure. Pain, it has been argued, is the representation of *tissue-damage*, and it is mediated by specific sensors devoted to this task.[85] Pain thus "tracks" tissue damage, even if I can experience pain without actually undergoing tissue-damage. Introspection, Tye argue, is a reliable process that takes awareness *of* qualities represented by the experiences as input and yields awareness *that* a certain kind of experience is present as output. This means that the concept of pain that we *apply* in the introspective act, may be purely phenomenal: our awareness of tissue-damage is thus mediated by other phenomenal qualities. That tissue damage is the quality paradigmatically represented by pains qua sensory experiences is an empirical hypothesis, not something supportable by a priori reflection upon concepts of introspection, Tye writes.

The affective dimension

Pain also has an affective, motivational, evaluative dimension, and it is to this we should turn if we are to find a suitable counterpart to pleasure. Pain, Tye points out, is normally very unpleasant[86]: we try to get rid of it, or to diminish it. We do this because it *feels* unpleasant or bad. The view that pain has distinct sensory and affective-emotional components was first proposed by Melzack and Casey in 1968 and has been supported by evidence ever since. Normally, both these components are present when we are in pain, but in some cases, the

[84] See the volyme edited by Aydede (2005).
[85] Melzack and Wall (1965), see Aydede (2002).
[86] Tye (2005).

affective component goes missing.[87] Pain is not *essentially* an aversive experience.[88] On the other hand, some very unpleasant experiences are not classified by their subjects as pains. Irritating itches are not sensorily classified as pains since the distinctive sensory content of pain is missing. But clearly, such unpleasant itches are part of the opposite of pleasure: it is this dimension, not the sensory classification, we need to account for.

While the experience of pain represents tissue-damage, Tye points out, it also represents it *as bad*. The affective dimension of pain is as much a part of the representational content of pain as the sensory dimension is.[89] This, of course, is where the problem starts. Representing something as bad in this sense, Tye argues, doesn't require concepts or any "higher" cognition, i.e. no full-fledged value concept, but is probably hard-wired from birth.

Aydede criticises Tye on the assumption that Tye defends a version of "strong representationalism".[90] Strong representationalism holds that the phenomenal content of en experience is completely exhausted by its representational content: to introspect such content is merely to have a thought about what the experience represents, as the output of the reliable process of representation. Even if we grant this for most perceptual experience, what can Tye mean by saying that pain experiences *represent tissue damage as bad*? How can that be the kind of property that can be detected or tracked? It is far from obvious what this property *is*.

Early representationalists, like Pitcher and Armstrong, argued that although pain experiences are genuinely perceptual, their affect is rather to be understood on the lines of a desire that the perception should cease. When in pain, the information about tissue damage is largely shadowed by this desire. Pain

[87] "Reactive dissociation", as Dennett (1978) calls it. See also Ryle (1969).
[88] See Hall "are pains necessarily unpleasant" (1989), and Stuart Rachel's "Is Unpleasantness intrinsic to unpleasant experiences" (2000).
[89] Tye (2005, p 107).
[90]Aydede (2005).

experiences are perceptual but also affective-emotional. And their affective phenomenology is not exhausted by their representational content. Barry Maund also points out the problem to account for the affective dimension of pain and pleasure as part of their representational content. [91]

Ned Block argues that the affective-emotional phenomenology of pain should rather be accounted for by a "functional role psychosemantics", whereas the sensation dimension can be accounted for by a more "informational" semantics along the representationalist lines.[92] The functional role of pain experiences is what gives it its particular (evaluative) content, which is then to be identified with the affective phenomenology of pain experiences. Rather than *representing* anything, we should simply say that playing this psychofunctional role *constitutes* the affective phenomenology of pain. With psychofunctionalism, we don't need representationalism, and besides: Motivation is not accomplished by representation alone.

Tye argues that the affective component is an *aspect* of the representational content: pain "feels bad". He suggests that pain represent badness *or aptness to harm*, and that this is an objective quality with which pain can be correlated. The way in which pain represents badness is similar to the bodily aspect of depression: one senses a departure from functional equilibrium. The shift in body landscape occurring as pain is experienced is not good for the subject: it is a departure for the worse, and this is what we experiences as bad. In this way, he writes, pain is usually an emotional experience as well as a sensory one.

A causal covariational account of the representational content of pain, including its affective character, says that an experience of pain represents location, tissue damage and aptness to harm. This representational content is nonconceptual, not just in the sense that the subject *need* not possess the concepts required to state the correctness conditions for the experience but that the content is of a

[91] Maund (2005).
[92] Block (2005, p 131-2).

kind that *could not* be the content of a thought or belief. But what, then, is there left for "representing" to mean? If "representing" just is being causally correlated, then there might be some truth to this account, but it does not seem to do anything to reduce or explain the nature of pain experiences. The essential fact about pain is still *how* that harm is represented.

1.4.3 A Representationalist Theory of Pleasure

If a similar account is to be offered for pleasure, what should it claim that pleasure represents? One suggestion, related to the suggestion that emotions are evaluative judgments, is that pleasure represents *goodness*. To assess this suggestion we must be able to say something further about what the good is. A desire-dependent view of the good seems suitable for this interpretation. The desire view just considered does suggest that pleasure could represent that our desires are fulfilled. Timothy Schroeder develops such an account in his 2004 book.

If the sight of something can be contrary to, or evidence for, a belief, visual experiences must have *content* in some way. Seeing, proverbially, *is* a form of belief. It is not as commonsensical that experiences of pleasure have content in this way. Pleasures are often referred to as "feelings" rather than "sensations", suggesting that they play a more subjective, self-reflexive role than that played by sensory perceptions. Furthermore, it is not straightforward that pleasure work as evidence for anything. But insofar as pleasure does have some evidential weight, it pertains to matters about what one wants and does not want. One consideration in favour of a representationalist view of pleasure is that we sometimes treat pleasures as something capable of being *justified.*[93] If pleasure represents anything, the most plausible candidate is that it represents something at least partly subjective. Like whether, and to what extent, our desires are

[93] See Perry (1967). Emotions like jealousy can be "justified", but pleasure on it's own can't be. Pleasure is nevertheless *part* of emotion.

satisfied. After due consideration, Schroeder suggests the following representational theory of what he calls the "hedonic tone":

> Representational Theory of Hedonic Tone (RTHT)2: To be pleased is (at least) to represent a net increase in desire satisfaction relative to expectation; to be displeased is to represent a net decrease in desire satisfaction relative to expectation. Intensity of pleasure or displeasure represents degree of change in desire satisfaction relative to expectations. (p 94)

This account, he points out, not only fits with our normal experiences of how pleasure work: it also makes sense of it. It explains why pleasure and desire are perceived to be intimately connected. It also explains why it is odd (but not unheard of) to experience pleasure and displeasure at the same time: the experiences say contradictory things. "Their contents are mutually exclusive". In depression, he adds, we can become "hedonically blind": we fail to experience pleasure because we fail to represent our net gain in desire satisfaction. There seem to be at least two ways of knowing that a desire is satisfied, and the situation for the depressed is like for one who cannot *see* certain colours, and yet believes them to be instantiated. In depression, the subject *misrepresents* the extent of his own desire satisfaction, Schroeder suggest.

The account also offers an explanation of what goes on in addiction. Uses of substances like heroin induce a representation of a net increase in desire satisfaction *when in fact no such increase exists*. Euphorigenic drugs "hijack" the brain's reward system.[94] While this makes sense, it is not clear whether such a diagnosis is open for Schroeder. Elsewhere, he argues that the reward system involved in hunger, for instance, has not food as its main objective, but rather a state of homeostasis. But if that is the case, the pleasures of heroin use might correctly represents the net increase of the satisfaction of *that* desire. In fact, many addictions seem to *change* our set of desires and preferences, so that the pleasures of drug use, sadly enough, might only too accurately reflect the state

[94] It is interesting to note that this representationalist view judge that euphorigenic drugs are actually a kind of hallucinogens!

of the agent.[95] This does not yet undermine the representationalist view of pleasure, however.

While it is true that certain drugs "hijack" the reward system, pleasure is, in fact, partly independent of that system: many addictions can be explained as people doing obsessively what they no longer get any pleasure out of. Even heroin does not seem to stimulate the pleasure centres *directly*, as it were, and is certainly subject to habituation and hedonic disappointment.[96] This means that the desire being satisfied by the drug use is *not* represented proportionally by the pleasures felt.

While intimately connected, the connection between pleasure and desire-satisfaction is not of the right kind for one to be *reductively* understood as representing the other. Pleasure might work as an *indication* of desire satisfaction, but we are still lacking an account of the nature of that pleasure.

Schroeder's view includes a specification of the influence of *expectations*. These are distinguished, broadly, into "intellectual" and "gut-level" expectations. Generally they go together, but they can come apart. Pleasure tends to side with the "gut-level", Schroeder thinks. This might be true in general, but there are "complications in real cases". Confident people tend to experience great pleasure at good news, and those of low self-esteem take bad news badly. This is so, Schroeder says, because it satisfies other desires, or fits into the picture of the self in a certain way. Experience sets a baseline of expectations of desire satisfaction against which new experiences are measured, which influence how they feel. Schroeder thinks that expectation is *decisive* for pleasure, but this seems too strong. Expectation tends to influence experience, but it is hardly decisive. It is simply not true that we only feel pleasure when our desires are satisfied to a greater extent than expected, and it is question begging if postulated at an unconscious level, even if it does seem to make sense of habituation. Some highly expected desire satisfactions might very well give rise

[95] See the work on addiction by Berridge (2003, 2004).
[96] See Berridge (2002).

to pleasure: whatever the relation, it is not proportional, and it is far from the decisive factor.

Now, Schroeder does not offer his theory as a *replacement* of the hedonic tone view, but as a specification of the content of experiences with this tone. He is not a strong representationalist: at no stage does he claim that *anything* representing desire-satisfaction in the specified way would thereby count as pleasure. It does not even suffice that we represent it *mentally*: I can represent, believe, judge any content to be true, and yet not experience any feelings. It seems that the content of perceptions and emotions is not exhausted by their propositional content. Nor is the nature of pleasure exhaustible in representational terms alone.

Leaving Schroeder behind, then, could we say that any *feeling* having this representational content would thereby count as a pleasure? If so, pleasure is the (phenomenologically heterogeneous) class of feelings that represents desire-satisfaction (or whatever). But by what powers does a feeling represent? If it is in virtue of some causal contingent relation, does that mean that any feeling whatever could have been pleasure? That seems unsatisfactory. If it is "in virtue of how it feels", we are back in need of a phenomenological account. Even if as a matter of fact pleasure is the only feeling having this representational content, we have yet to capture what this feeling *is*, and this seems impossible in purely representational terms.[97] In addition, it is questionable whether representation can be an intrinsic feature of an experience or, indeed, of anything, which means that the same worries arise here as for the extrinsic desire view on pleasure. In fact, on the proposal considered, representationalism is a *version* of that view, with the qualification that pleasure is not the object of desire, but the representation of desire satisfaction.

[97] Ledoux (1996) point out that one of the differences between feelings and "mere thoughts" is that, first, they are partly generated by different systems in the brain but more importantly: feelings involve many *more* brain systems than mere thoughts.

While pleasure might very well represent something, this is not the essence of what it does: to the extent that all pleasures do represent something, they do so by having some *other* property in common which carries that content, or performs that function. As the neuroscientist Kent Berridge puts it: Emotional reactions typically involve extensive cognitive processing (…) but emotional processes must also always involve an aspect of *affect*, the psychological quality of being good or bad".[98]

[98] Berridge (2003).

1.5 The "Adverbial" View

While representationalism has something to it - pleasure has some sort of informational role - it fails to provide a reductive basis for a theory of the nature of pleasure and displeasure. The affective character of emotional experience is not reducible to *what* is represented. In order for a mental state to be one of pleasure or displeasure, it is essential *how* it is represented as well. Guy Douglas argues that "I feel pain" is an answer to the question *how* do you feel, not *what* do you feel.[99] This claim is typical for what is sometimes called the *adverbial* view.[100] What makes an experience a pleasure is not what you experience, but *how* you experience something.[101]

Pleasure, according to the adverbial view, is a mental *state*, rather than a mental *object.*[102] While these states might represent something, that is not the essence of their kind. Pleasures also typically cause behaviour, and probably often do so on the basis of what they represent, but again, that is not their essence. Offering a view of this kind, Karl Duncker argued that pleasure is an *incomplete* feeling.[103] Pleasures always qualify some other experience; it is a hedonic *tone*. Rather awkwardly, Duncker reserves the name 'pleasure' for this tone, rather than for the experience to which it pervades.

The pleasure dimension

A different way to formulate this "aspect" approach to pleasure is to speak, as Alston does, of a pleasure-displeasure dimension on which particular

[99] Douglas (1998).
[100] Gosling (1969) treats "adverbial", not as much as a form of having an experience, but rather as "willingly, with desire . In his view, Ryle (1969, 2000) counts as an adverbial theory of pleasure
[101] The implication being that this "how" is not reducible to just yet more information.
[102] Moore (1993) argued that since one can be conscious of ones pleasure, pleasure must be distinct from that consciousness, but this view suggest that Moore get the categorisation wrong: the pleasure is in the mode of consciousness, not in its object.
[103] Duncker (1941).

experiences may vary.[104] It is the intensity of this dimension, and not the intensity of the sensation to which it may attach, that matters. Alston points out that there is a *binding* problem for experiential qualities in general, which is highly relevant for states of pleasure, who are often intimately connected to the activity we take pleasure in. It is clearly not just that they appear simultaneously in the same consciousness. We must posit a more intimate connection between pleasure and its object and it seems impossible to specify such a bond if we interpret pleasure as a kind of sensation.[105] If, on the other hand, we say that pleasantness is a property that a sensation can have as one of its qualities, the binding seems to be implied: the property of pleasantness belongs to the same experience as the sensation, say, to which it attaches.

Kagan proposes that we use "volume" as an example of the kind qualitative "dimension" pleasure might be.[106] Volume is a property that essentially belongs to sounds, and one cannot imagine it occurring "on its own". Yet there is something it is "like" to experience loud sounds. The heterogeneity argument for pleasures may be repeated for volumes: loud sounds do not all sound alike, but that doesn't undermine the "distinctive feel" of volumes: there is *something it is like* to hear a loud sound, they form a kind, distinguished by how they feel. We can defend the feeling view of pleasure along the same lines. If pleasure is a dimension, rather than a sensation or a component of a sensation, it couldn't be had in, so to speak, *isolation*. Neither volume nor pleasantness is a *component* of the experience. Even so, some very loud sounds have negligible other components: their loudness is their most distinct feature. Equally, some pleasant experiences are first and foremost pleasant, and they might in fact be cases at least bordering on "pure" pleasantness. As Crisp points out, the distinction between "dimensions" and "components" is spurious.[107] Loud sounds form a kind, after all, so why not say that they are a component of the

[104] Rachels (2004) argue that the antonym to pleasure is displeasure, not pain.
[105] Alston (1967, p 342).
[106] Kagan (1992).
[107] Crisp (2006).

experience?[108] The distinction depends on your meta-physics of parts and wholes when it comes to experiences.

A clear benefit of the adverbial, dimensional view is that it recognises that experiences may be *complex*, exhibiting a variety of aspects.[109] This means that it can, to some extent, *accommodate* heterogeneity: pleasures feel differently because the hedonic dimension latch onto experiences that in other respects can vary as much as you like. But it cannot accommodate "radical" heterogeneity. The argument, after all, was formulated as if we find *no* experiential quality in common between the experiences we call pleasant. If anyone keeps insisting that there is in fact *nothing* these experiences have in common, as to how they feel, we can only say that he is missing out, or use the term in a different way from us. But lets also note that not all dimensions, and not all tendencies we have to group experiences together, need be noted.

If this is what the adverbial view comes down to, it is, as Sobel argues, not as much an alternative to as a *version* of the feeling view.[110]

[108] The colour-analogy appears as early as in Plato's Philebus: that pleasures are alike as colour is to colour, but that black and white are still opposites, Socrates points out.
[109] Aydede offers a similar argument for pains (2000).
[110] Kahneman, Waker and Sarin (1997) calls it an "attribute" of an experience.

1.6 Pleasures are Internally Liked Experiences

Pleasures, I believe, should be understood as Internally Liked Experiences. They are experiences partly constituted by an attitude which itself is experienced. This view is distinct from the desire-oriented view that claims that pleasures are experiences that are merely the objects of, or otherwise externally related to, a pro-attitude. It is thus an internalist conception of pleasure. I believe that the relevant attitude is part of the pleasure experience. However, I also believe that the pleasure is usefully understood as the *object* of that attitude.[111] This may sound awkward, or even viciously circular, but it is important to understand what kind of claim this is: it is a way to describe an experience in terms that were not developed for it. The object/attitude distinction is not as obvious when you are dealing with experiential properties, as it is in the archetypical desire-object relation familiar from the literature on propositional attitudes. Some experiences are complex, i.e.. they are units consisting of a number of experienced aspects where each aspect can be singled out for attention. If one aspect of the experience is attitudinal, it may take the *other* aspects, or the whole of the experience, as its "object". What is liked is how a certain experience feels and part of how it feels is how this attitude feels. It may, in fact, be that very thing that we like about it. This "circularity" is no different from when, say, I like my life and part of my life is that I like it.

Now, it may be objected that requiring that the experience be the *object* of the attitude is unnecessary.[112] Surely, if pleasure is the experience of liking something, this experienced liking may take *any* object. In particular, in the cases that the pleasant experience has some external object, why not say that the liking to has *that* as its object, rather than the experience itself?

[111] I argued for this in Bengtsson (2003, 2004).
[112] I'm indebted to Jens Johansson for pointing this out to me.

I agree that pleasures that take external objects may very well be said to be likings of those objects. The argument from misattribution seems to lend further support to the introspective veracity of that claim. In addition, what is truly important to the account is that the attitude is experienced, and that this is a *part* of the pleasure experience. This being said, pleasures are properly viewed as essentially positive *in themselves*: the experience you have when you are experiencing pleasure is itself something that is being liked (even though this need not be noticed by the agent). The difference between merely liking something and taking pleasure in it, on this view, is a matter of how it feels, and this, I believe, is best captured with the statement that this feeling itself in the latter case is the object, or rather *an* object, of the attitude in question. The experience is, at least, the *proximate object* of that attitude.

1.6.1 Simply Feeling Good

Phenomenal experiences (qualia) have often been understood as *simple* entities or events, examples of which typically provide a single experienced property, like the sensation of red. In this fashion, pleasure, as Leonard Katz points out in his impressive, nearly book-length entry in the Stanford Encyclopaedia of Philosophy, has been conceived of as a "simple uniform feature of momentary conscious experience, that is *obviously* good in itself and consequently attractive to whoever experiences it".[113] This formulation brings something important to the fore: pleasure feels *good*. Proposed as a mere synonym to 'pleasure', this might not say much, but it does seem to capture something important about the nature of pleasure; both about how it feels and about its function in human psychology. Indeed, as I will argue in part 2, seeing how the analysis of evaluative judgments is controversial and notoriously difficult to get right, it doesn't provide much illumination of the nature of pleasure to say that it has a related content to those judgments. Rather, and that is the main point of my

[113] See Goldstein (1989) on the intrinsic value of pleasure: pleasure is valuable because of its intrinsic features.

argument; it is the other way around. Feeling good is the epistemologically and ontologically "prior" evaluative phenomenon. Goodness, I'll argue, is primarily an experiential property.

The simple, singular view of pleasures fits badly with introspective evidence, and in particular with the intuitively compelling claims that pleasures are essentially (or even just potentially) *incomplete* as experiences. They are always (or potentially) the pleasure *of* something else, and the binding between a pleasure and its object is stronger than mere simultaneity would guarantee.[114] This is the point to bring home from the argument for the "adverbial" view. While we do not always distinguish aspects of our experiences, it seems clear on closer inspection that aspects of an experience can fluctuate independently and the experience still remain the same entity. This is not to say, yet, that pleasure could *not* appear on its own: Whether or not there are such things as *pure* pleasures, an experience can be *primarily* pleasant, and have negligible other properties. Insofar as the "uncompleteness" argument tracks any truth, it is that pleasures are practically always *triggered* by some other experience. Pure pleasures could be artificially induced, via electric or chemical stimulation. When this is done, we do tend to associate that pleasure with whatever else is going on in consciousness at the time, but there is, I to my knowledge, no absolute obstacle for purely pleasant experiences.

1.6.2 The Truth in Desire-theory

There is something irreducibly positive about pleasure. This makes an attitudinal, desire-oriented theory of pleasure plausible: an undoubtedly positive element is given a definitive role in the definition of pleasure.[115] It also explains

[114] Nevertheless, simultaneity could *cause* experiences to "bond" in the intended sense, due to so called "Hebbian learning". See Ledoux (1996).
[115] Sidgwick (1981).Katz (1986), Gosling (1969), Perry (1967), Schroeder (2004) offer very similar arguments. Gosling points out that (p 154) what kind of sensations and bodily, visceral states (excitement or relaxation etc.) are positive depends on *temperament*, and, most importantly, on what the subjects *likes*.

why we take pleasure to be a *reason*, as something to pursue, and it also offers a *causal* account of this pursuit: pleasures are pursued because, by definition, they are objects of desires. The central claim of the desire-oriented view, I take it, is that a pro-attitude, be it general or specific, *makes* an experience pleasant. But what does this "making" involve? The dominating idea in the desire-theory is that this "making" relation states nothing beyond the fact that this particular experience is the object of a pro-attitude. Similar attitudes can be taken up towards any object or state of affairs: it's only because of the accidental fact that this attitude takes an experience as its object that the objective come to be called a pleasure. Alternatively, and, as I think, more plausibly: the attitude actually has an *impact* on the experience. This might suggest that a desire *causes* an experience to be pleasant, but that being pleasant is ontologically independent from this cause. In that case, the desire would seem to be irrelevant to the essence of pleasure, and we would have all but abandoned the desire theory. But there is yet another "making" relation to be considered, namely the view that the relation between the attitude and the experience is not accidental, but *constitutive*: the attitude is a constitutive part of the pleasure experience.

In the section treating the desire view, we considered the possibility that the relevant desire be not a disposition, but a feeling. This can now be put to use. If the relevant sense of desire is a form of feeling, why should we take pleasure to be the experience *desired* rather than that feeling of desire itself? The experience seems to be just an occasion or cause of the desire which itself is the decisive feeling. If the experience of the attitude is what gives the event its distinct experiential character it doesn't matter whether or not any other experience is the "proper object" of that attitude: when you like something, experiences of it, or just related to it, tend to change, they are assigned importance and become worth attending to. Ultimately, some of these experiences may get *pleasant.* The relevant attitude, an attitude I call "liking" is *part* of the experience, when it is in this sense liked. It is part of it, because it not only attaches to it, but modifies it. It is not merely simultaneous to the experience.

The term 'liking' is preferable to 'desired' or 'wanted', since the latter two seems more properly assigned to dispositional states, whereas 'liking' is directed at occurrent objects. It is possible to want something that you do not like, and vice versa. In the end, what's important is not, strictly speaking, that you get what you want, but that you like what you get.[116] The external desire connection is not tight enough to make sense of the positiveness of pleasure.

The attitudinal reading seems fitting for cases like pleasant sensations, but we also speak of being in pleasant "moods", i.e. in states without any particular object. Leonard Katz suggests that pleasure might instead be understood as a *stance* of "affective openness", welcoming or immediate liking.[117] While paradigmatic attitudes like beliefs and desires are individuated by their propositional or "property-self-attributed" contents, instances of such a stance can be individuated via *intrinsic* features, more like "stuff or process" than as particular mental acts. Pleasure on this understanding is not an object-bound attitude, but its *own* thing, even when divorced from content-directed thought and motivation. This seems to fit with the kind of state we find ourselves in pleasant repose or in meditation.[118] The view I propose is certainly intended to cover such states.

1.6.3 Explaining Heterogeneity: Complex Phenomenology

Pleasures are a set of experiences, distinguished by how they feel. The view I propose does not differ much from the "standard" view in that respect. But the heterogeneity argument is not wrong: pleasures do feel different from each other. The key to this seemingly contradictory statement is that pleasures are at least potentially *complex* experiences. They are heterogeneous because they can vary in all the *other* respects, in all other felt aspects of the experience. What this

[116] Fred Feldman express his view (2004) in very similar terms.
[117] Katz (2006), James Russell (2003) one of the pioneers in so called "happiness research" speaks of "in-itself objectless feeling good" at the ground level of the construction of more complex positive emotions.
[118] Katz (2005).

means is that pleasures are not *entirely* heterogeneous, and thus the claim is not consistent with a stronger heterogeneity claim. The pleasure of an experience need not be the focus of *attention* in a pleasure experience, but might be present in a background capacity.[119] This opens up a number of questions: what does it take to be a pleasure? Is it enough if the experience includes only a minuscule amount of this felt affective value? This seems to me to be a matter of little importance: normally, perhaps, when I say that I see red, I mean that there is something notable that is red, but there are certainly situations where the red that I see, and report, is a very negligible part of my experience indeed. It might be part of a test of my eyesight, for instance, or a "spot the red spot" competition. Similarly with pleasures: When asked whether I find a sensation pleasant, I may very well answer in the affirmative, even though the pleasure I feel is very small indeed. Sometimes when asked what we experience, however, we are asked about what *dominates* our experience. We could of course postulate that an experience is a pleasure only if its pleasantness makes up more than 50% of the experience, but it is hard to make sense of what that would mean. A more difficult question is what we would say about an experience with a note of pleasantness that is nevertheless predominantly *un*pleasant. Would it be a pleasure? Whereas pleasures and pains are often construed as opposites, the view proposed seems to allow that one and the same experience incorporates elements of both, and indeed, this seems to be at least one aspect of what being a masochist is all about. We can certainly enjoy pain *sensations*, but can we enjoy *unpleasant* experiences? On my view, there are no conceptual reasons to think it impossible.[120] In general, it is possible to experience opposites simultaneously, as when we feel both hot and cold at the same time.

[119] Aydede, (2000) Crisp mentions it (2006), and there are elements suggesting this argument as early as in Epicuros, and in Locke (1975). Alston (1967) mentions that "feeling theories" of pleasure can say that the difference between pleasure consists in what bodily sensation is involved: what makes it a *feeling* depends entirely on the quality on the pleasantness-unpleasantness dimension.

[120] There are at least some empirical reasons to believe that they don't naturally occur that way. (See Schroeder (2004) and Katz (2006).

1.6.5 Evidence From the Affective Sciences

Treading very carefully, because these are dangerous grounds for a non-expert, I venture to claim that evidence from the affective sciences with regard to pleasure, affect and its place in human nature does support the view considered here.[121] In particular, there is evidence that a there is a distinction between two types of pro-attitudes, where one is more dispositional, action conducive, and the other experiential.[122] The neuroscientist Kent Berridge calls them "Wanting" and "Liking". Wanting and liking are two different desire like/hedonic states/processes in the brain that often occur together: they are part of the "dopamine-opioid hedonic circuit". Dopamine is the neurotransmitter most associated with motivation and drive, whereas the opioid system is associated with experiences of pleasure. Their interconnectedness means, in commonsensical terms, that you tend to like what you want, and want what you like. While intimately connected, these two systems are functionally and anatomically distinct, and they can come apart.[123] This is of course quite common, as we often find that even if we get what we want, we might not like it. This conclusion was reached by the desire-theorists on independent, theoretical grounds, but the evidence suggest that rather than a particular desire-object relation being relevant, there is a distinct kind of attitude, liking, associated with hedonic experience. More worryingly, the two systems can become more radically dissociated. Many addictive behaviours can be explained as cases where we keep on "wanting" what we get no more pleasure out of, or where the hedonic reward is not worth the effort.[124]

Pleasure and reward

The operational, functional definitions of pleasure tend to focus on its status as unconditioned reward: where a reward is defined as something for which the

[121] For a more informed overview, see Katz (2006).
[122] See for instance Kahneman et. al. (1997).
[123] Berridge (2003, 2004, 2007).
[124] Berridge (2003), Kringelbach (2009).

subject is willing to work.[125] The distinction between wanting and liking can be put in relation to this notion of reward: when a conditioned reward attains self-sufficiency, we can be said to want something that we get, and yet we need not like it. There are both advantages and disadvantages in this arrangement. If a conditioned reward looses all connection to pleasure, we can be described as obsessed, and we loose the ability to unlearn behaviour.[126] A healthy disposition requires that we be able to disengage with activities that have ceased to be genuinely rewarding. On the other hand, success in life also seems to require that we can occasionally forgo direct reward in order to attain other goals we may have, goals that may, indirectly, be the route to a more rewarding outcome in hedonic terms.

Unfelt rewards, unfelt pleasure

Berridge points out that reward in the functional sense, i.e.. as an effective cause of behaviour and learning may be imperceptible.[127] An effective reward may take place below the "threshold" of consciousness. He argues that whether to treat such a state as a pleasure or not, i.e. whether consciousness is essential to pleasure, is a matter of semantic taste.[128] One could make a case that this functional sense of reward is not sufficient for the "common sense" term, but "reward" is intended as a technical term and should be kept as such. It is not difficult to make sense of the phenomena of "unfelt" pleasure while starting out with an experiential concept. We can truthfully report being pleased with progress, say, or in love, without necessarily feeling anything during that report, and yet the truth of those statements is based on the occurrence of felt pro-attitudes. I may still *have* the attitude, just not occurring in my stream of consciousness at the time. It is like belief, in that respect. The functional sense

[125] Berridge (2004). Whereas some hedonic theory of unconditioned reward is very likely true, it is a complicated story See for instance Wolfram Schultz (2000) and Berridge (2007). Schroeder (2004) questions the reward theory of pleasure on the grounds that rewards can *cause* pleasure. See also Davidson et. al. (2002).
[126] Kringelbach (2009).
[127] Berridge (2003).
[128] Unfelt affective reactions in Berridge (2002, 2004).

can occur below he threshold of consciousness, and yet the conscious phenomena be essential to the category.

1.6.5 Pleasure and Content

Pleasures do seem to have content, and to share the content of a thought. But pleasure is not reducible to mere thought, and it is not analysable in terms of belief or judgment alone. As for its content, the most plausible content proposed is that hedonic experiences are forms of *evaluative* judgments. Nevertheless, our pleasures may occasionally go *against* our evaluative belief or judgment, which is much the same phenomenon as the visual illusions of how a stick believed to be straight, still looks bent if partly submerged in water. Aristotle, for one, bought into this idea and regarded pleasure as presenting a fallible appearance of goodness, which might differ from our rational belief. The cognitive evaluation or appraisal postulated by psychologists is acknowledged as a fast, automatic bit of neural information processing, attainable even by creatures with weak, if any, conceptual abilities.

In order to have the same content as a thought, which pleasures *may*, but need not have, they would seem to have to have *propositional* contents. We normally say that we enjoy, or take pleasure in, things other than experiences, like states of affairs, i.e. objects of a propositional form.[129] Feldman, who believes that all pleasures take this form, argues that the key phenomenon is the attitude "taking pleasure in".[130] This makes it possible to say that pleasures are true or false, namely if their propositional object is. More importantly, we can even say that some pleasures are *bad.*[131] Since there is an attitude and an object, we can ask whether the attitude *fits* with the object, and hold us accountable for the

[129] Whereas Chisholm (1986) thought these should be accounted for as a version of sensory pleasure, Feldman (1997a) and Heathwood (2006) thinks it's the other way around.
[130] Feldman (1997a) takes this stance as a primitive, as does Heathwood (2006). Feldman believe that the pleasure is the states of affairs consisting in this attitude and its propositional object. This makes it possible for him to let the value of the pleasure depend only on intrinsic features of the pleasure.
[131] Lemos (1994) drawing on Chisholm's take on Brentano (1986), See also Zimmerman (1989).

attitude. This is useful if we would like to distinguish between good and bad pleasures. If the pleasure is the object of the attitude, the pleasure itself is just a plain fact, but if it is the *attitude,* it can be assessed in accordance with how it fits its object, i.e. if the object is worthy of such appreciation.[132] While this is a very clever solution for hedonists, I believe, as will be argued extensively in the next part, that this gets it the wrong way around. The problem of how experiences of pleasure can have the "same content" as a propositional judgment is not the problem how to make pleasure more like a propositional attitude. The *primary* value phenomenon, I argue, is the experience of pleasure, and the problem is how a *judgment* can have this kind of content.

1.6.6 Internal Likings

I propose that pleasure is a phenomenological kind: what pleasure has in common is how they feel. Their distinctive feel is usefully thought of as *attitudinal*, it is the experienced *liking*, or even *evaluation* of something.[133] This fits with the universal perception of pleasure as something essentially *positive*. In fact, it is positive in two senses: it *is* a positive evaluation *and* the object of this positive evaluation. Of course, the fact that pleasure essentially involves a positive attitude does not stop it from occasionally being rejected and avoided: something that is internally liked can be externally disliked. Indeed, since the view proposes that experiences may involve contrary (but not contradictory) elements, one and the same experience may be both internally liked and disliked at the same time. The recognition of experiences as complex events also allows us to account for the plausibility of the heterogeneity argument, without having to acknowledge that pleasure and displeasure "feels the same". Admittedly, we are denying a strong version of the heterogeneity claim, but that seems to be just as well. In deference of a strong version of that claim, we can grant that there are experiences that are intrinsically (but externally) desired

[132] If the attitude is towards the pleasure itself, Lemos argues, the object of the attitude seems to be morally neutral, and there is at least nothing objectionable about such a pleasure that would undermine its goodness. Lemos (1994).
[133] Pleasures, in terms borrowed from Helm (2002) are *felt evaluations*.

because of how they feel which nevertheless do *not* have this felt quality. Whether or not to call them pleasures is perhaps up for grabs, but it just seems that, first, calling this class of intrinsically desired experiences pleasure despite their lack of the felt quality misses the fact that there is such a thing as the feeling of pleasure. And, second, it misses the fact that this feeling is often what we desire an experience intrinsically for.

Even if what *makes* an experience a pleasure is somehow a "simple" idea or quality, this doesn't mean that pleasures have to be simple, isolable experiences. Experiences classified as pleasures are frequently quite complex, and they vary according to what else is included in the experience. I'm suggesting that the quality *makes* an experience pleasant, but that the word "pleasure" names the experiences thus qualified, not the quality itself. This is just a matter of convenience: it seems to fit better with everyday talk.

By making the attitude internal to the experience, we also make sure that what makes an experience a pleasure is internal to that experience. This means that the experience of pleasure can have intrinsic value. A further advantage of this theory is that, by noting its attitudinal nature, it makes sense of the *subjectivity* and *individuality* of pleasure. An auditory experience, say, that is pleasant for me might not be so for you, and the difference is to be found in our respective attitude towards the experience: I like it, and you don't. But this "liking" permeates the experience: my experience is a pleasure, and yours isn't.[134] But saying that the difference is that I like it and you don't also say something about the causal precursors of our differences. The *reason* why I like it, in the hedonic sense, often has to do with my previous desires, interests, beliefs, activities.[135] Because of the connections that hold between higher cognitive thoughts and attitudes and more basic affective processes, what makes for the difference of

[134] People differ in their imaginative capacities, so that some people just can't understand what it would be to experience this taste, say, or that sound *pleasantly*. People who can imagine, however, are imagining an experience that is slightly different from the one they typically have in that situation.
[135] As Helm points out in his (2001) and (2002).

experience may very well be the beliefs and desires I hold about the thing experienced and you don't.[136] And in the other direction, the type of experienced liking I have had the privilege to experience frequently influences those higher cognitive attitudes and habits as well. What I experience when I like some experience is *partly different* from what someone experience who does *not* like it. The chief difference is the experience of liking, but it is not *only* that. Often, I take it, the one who likes something will focus on different features of an experience or event than the one disliking it or the one being neutral towards it. It is also often the case that thus liking something makes you take pleasure in many other experiences related to that thing as well. In the long run, as we shall see in the next part of this book, this fact about pleasure accounts for the plausibility of relativism: what is good, i.e. what is pleasant, depends on our attitudes. But the bearer of that value, that state of pleasure, is the same thing for all. The occurrence of a subjective state is still very much an objective fact.

[136] Aydede (2000).

Part 2: Value

2.1 The Theory of Value

> *It is the nature of a hypothesis, when once a man has conceived it, that it assimilates everything to itself, as proper nourishment; and, from the first moment of your begetting it, it generally grows the stronger by everything you see, hear, read, or understand. This is of great use.*
>
> Laurence Sterne, Tristram Shandy

2.1.1 Fundamental questions

In this, the slightly more daring and ambitious part of the thesis, I aim to define and defend a version of naturalistic hedonism about value. This theory claims that there are value facts, that they are natural facts, and that they are "hedonic" in nature. The pleasures painstakingly defined in the first part will, in other words, now be put to use. But before doing so, there are some questions that need to be raised. What is the subject matter addressed by value theory? What is a theory of value supposed to do? Defending a particular theory of value, I might be expected to be able to answer these questions. As so often in philosophy, though, specifying what's at issue is part of what makes the problem so difficult. Specifying what's at issue is precisely what *is* at issue.

In order to provide a plausible account of value, we need to engage with questions about metaphysics, epistemology, semantics and psychology. These are the philosophical foundations on which meta-ethical theories are built, and from which they lend support and argument. These philosophical disciplines are included in the wide set of considerations that bears on value theory. But we

will also have to raise questions about the nature of theory and theoretical considerations in general, and how they apply to this subject.

It is important to understand what this project involves: What I ultimately want to accomplish is to argue for a particular theory of value. In order to do so, however, I need to establish what it is for a theory to be a theory of value at all. Throughout the next few chapters, I will make a number of claims on this issue in order to secure that the version of hedonism I defend qualifies as such a theory. This, I believe, is the critical point in the argument.

In order to get closer to our subject matter we'll take a look at the diverse *options* available in meta-ethics and value-theory. I will make a point of the rather fundamental disagreements that pervade this discipline, but the point will not be a sceptical one. Rather, taking stock of the various alternatives held worthy of consideration will help us find out what sorts of arguments and evidence hold currency in this domain.

The point of noting fundamental disagreements is to justify the theoretical "stretch" necessary to defend a theory with some amount of specificity. Since more or less everything in value theory is up for discussion, specific claims are bound to rub some people the wrong way. Paying attention to the range of positions held worthy of consideration in meta-ethics is a way to justify the theoretical stretch necessary to achieve such specificity: If controversial decisions are inevitable, they no longer represent a singularly theoretical cost. In so far as observations about the nature of theory and the variety of available candidates support scepticism/pessimism, they do so regarding the outlook for finding a single acceptable standard for a correct theory of value. Rather than justifying the project all the way, then, we can argue that *given* a certain conception of what a theory of value is supposed to do, an interesting and ultimately true version of the preferred theory can be construed. This weaker claim in favour of hedonism is the minimal result of the theory developed in the next few chapters.

2.1.2 The subject matter and nature of value theory

Trivially, the subject matter of value theory is "value", or "the good".[137] To say this is obviously to say very little. If anything, it is to say that the theory has something to do with value *judgments*: their truth, perhaps, or what they refer to, or their justification, or what they might mean. Value theory is at least partly constrained by what *we* take to be valuable, i.e. by our *substantive* values. While we might be willing to revise our value judgments in the light of some considerations, these judgments can still be used as a starting point for a theory of value.[138] Arguably, a theory that did not somehow latch onto our actual evaluations would not be a theory of value at all. First-order value judgments belong to our most obvious data in this domain. To what extent, and in what way, the content of those judgments constrain and determine value-theory, and vice versa, is yet something to consider.[139]

We need to address, too, the question what kind of theory value theory is, and what we can expect from it. Should it provide a *conceptual analysis* of 'value', i.e. extricate the meaning of evaluative terms and statements? Or do we expect it to pick out a referent for evaluative terms, a property (or properties) capable of making evaluative judgements come out as straightforwardly true? Should the theory provide us with a method to vindicate evaluative judgements? Which of these questions a theory must answer in order to be a proper theory of *value*, and *how* it must answer them, is unclear. I take it that a theory of value is a theory that addresses any subset of these, and perhaps other, related, questions.

To address all of these questions, and to do so in a unified manner, would be desirable. If such an account were available, it would be eminently eligible as

[137] Chapter 1 of Moore's Principica Ethica (1993), "the subject matter of ethics". I will treat these two terms as synonyms.
[138] Realising ones fallibility might even be a competence-requirement for evaluative terms.
[139] Importantly, no *particular* answer to this question seems required Rawls (1971), Brandt (1985), Daniels (1979), Tersman (1993).

our theory of value. If such a unified account cannot be presented, however, we face some hard theoretical decisions. How much does a theory need to account for in order to cover *enough* of the value relevant issues?

Definitions

In what is arguably the starting point for modern meta-ethics, G.E. Moore (1993) argued that meta-ethics is concerned with the *definition* of 'value'. Ethics, he argued, would be effectively useless if an acceptable definition of this term is not given. It is only in light of a definition of our subject matter that we can decide what counts as evidence and justification for ethical/evaluative judgments. But Moore also pointed out that in the *preliminary* stage of ethical theory, presupposing any particular definition of 'value' would alienate people with whom we are properly thought to substantially disagree. We should start out allowing, that as far as the commonly known meaning of 'good' goes, anything could be good (p72).[140] Moore then famously concluded that goodness is a simple, unanalysable property. To avoid constraining its applicability, Moore stripped 'value' of descriptive content. The force of this latter argument is disputable, but the problem is not: we do view ourselves as being in a proper disagreement with others over what's valuable, but we do not want to say that their notion of value is different from ours. We construe it as *disagreement*, after all. There must be a common subject matter over which we are disagreeing.

The argument for definitions, if successful, seems to apply quite generally: what is acceptable as evidence in *any* discipline presupposes 'definitions' of a subject matter. So what are the definitions on the basis of which something can be used as evidence in *meta-ethics*? What is it that theories properly construed as meta-ethical have in common? There is a problem with transporting the argument in its entirety to meta-ethics: we could hardly say that as far as the commonly

[140] A very similar argument is made by Ewing (1939), who wants to rule out any definition of "good" that only allows for experiences to be good. It might be true that only experiences are good, but it is not conceptually true.

known meaning of 'good' goes, it could *mean* anything whatever. How, then, can we avoid alienating meta-ethicists with whom we are properly thought to disagree? What is *our* common subject matter? When arguing about definitions, we cannot presuppose those very definitions; we need to take a step back in order to find the support we need.

John Mackie, in another central text of 20th century meta-ethics, took his subject matter to be the meaning of 'good', understood as "the most general term" in ethics.[141] Whereas there are a number of ways in which the *word* 'good' can be understood and used, the meaning of *this* term, he argued, does not change with context. The *extension* of the term can change with context, but something distinctive must be held constant over these varying uses for them to be recognisable as variations. There must be some *core evaluative meaning* to the various expressions that inhabit our evaluative language.

Now: this common element might be too thin to correspond to any of our everyday concepts.[142] Familiar evaluative concepts might be semantically *thicker* notions that are only partly constituted by this core notion. Nevertheless, it is this common element we need to isolate in order to make sense of evaluative notions. In this book, the proposal is made that pleasure provides that element.

2.1.3 The primacy of semantics, the analytic and the a priori

It can be argued that in order to assess whether there is a metaphysical question about value, and whether evaluative statements require epistemic justification, we need to answer the question about the meaning of those statements and terms. If value judgments and terms are not descriptive or attributive there seems to be little sense, and less point, in asking the metaphysical and epistemological questions.

---

[141] Mackie (1977).

[142] Moore in fact wrote that the simple property "goodness" was such that, "all the moral words refer to it"; he did *not* say that any moral word *as actually used* was synonymous to it. I don't offer this as an interpretation of Moore, however.

Taking the importance of these questions to imply that meta-ethics should primarily be concerned with philosophy of language, however, would be to rely on an artefact from the 'linguistic turn' in meta-ethics nowadays largely viewed with suspicion.[143] For one thing: There exists no uncontroversial *and informative* analysis of the meaning of evaluative statements. This would seem to imply that the semantic approach, if understood narrowly as somehow independent from, or prior to, other areas of theoretical investigation, is a dead end. Semantics understood as the a priori analysis of terms can only help us when we have a clear notion of what type of concept we are dealing with. It could then help us determine the application of that concept in particular cases. It can*not* help us when this meaning *itself* is under scrutiny. Other input is needed for a theory to get of the ground, and theoretical virtues other than linguistic accuracy are needed in order to make a convincing case in meta-ethics. Facing conflicting analyses, and lacking a neutral way of settling the conflict, we cannot hope to reach agreement by appealing to analytical facts alone.[144]

Value theory should be as much concerned with the state expressed and the world encountered in evaluation, as it is with the meaning of terms used in evaluative discourse. Of course, if we conceive of semantics broadly enough, those metaphysical, epistemological and psychological matters might be fitted into it. I suppose it is this tendency to treat semantic as covering more or less *all* of philosophy that accounts for the plausibility of the linguistic turn.

If we believe, with the semantic externalist, that the meaning of terms depends on what goes on in the world, the quest for the subject matter of value theory depends on whether there exist suitable properties for evaluative terms to refer to.[145] If it turns out that our evaluations are causally regulated or responsive to some particular natural property, detecting or reacting to that property is at

[143] Bernard Williams points this out in "Ethics and the limits of philosophy" (2006). See also David Copp's introduction to "Morality, Reason and Truth" (1985).
[144] Rawls (1971) points out that there are virtually no definitional a priori truths in moral theory.
[145] See for instance Putnam (1973) .

least part of what evaluative states do. If there is such a property, and such a relation, this fact should be present in any comprehensive theory of value, even if it need not be part of the analysis of the term, or thought to be essential to the property of value itself.

The point of these observations is that arguments trading on a particular view of the meaning of 'value' are insufficient to rule any conflicting theory out of consideration as a meta-ethical theory. In particular, no hedonist worth his/her salt should be discouraged by the claim, or even required to deny, that 'value' does not mean 'pleasure'. Pleasure, according to the version of hedonism defended here, should primarily be understood as what value *is*, not what it *means*. Even so, the theory offers a possibility to provide the "core evaluative meaning" sought by Mackie.

The theory, then, is not primarily a theory about the meaning of value-statements, or the contents of evaluative concepts. It is, rather, a theory about the *property* of value. It is a version of what has been called *metaphysical* naturalism. This could be understood in relation to Alan Gibbard's claim that value might be a natural property *even though* 'value' is not a natural kind concept.[146] Value claims are true in virtue of natural properties, but to say that something is valuable might not be merely to predicate this property to it. If true, this would mean there is something *lacking* from the naturalist account of value, and I will propose that nothing of importance is. Nevertheless, I'm no stranger to the idea that terms can have multiple uses, and that this tends to influence their perceived semantics.

The approach to the problem of value needs to be comprehensive.[147] The point is that the nature of value, if it has one, and even *whether* it has one, could be

[146] Gibbard (2003), and in "Normative properties" (2006).
[147] See chapter 2.4. Putnam (1981) pointed out that "...it takes empirical and theoretical research, not linguistic analysis, to find out what temperature is (and, some philosopher might suggest, what *goodness* is), not just reflection on meanings." (p 207).

investigated by methods other than conceptual/ linguistic analysis understood narrowly.

2.1.4 Surface grammar and function

The surface grammar of evaluative language suggests that 'value' is a property name: things are regularly characterized as being good or bad. Should this be taken at face value? One of the reasons for doubting the surface grammar of evaluative judgments is that no property seems to have what it takes: no property is such that to ascribe it could intelligibly be all we do when we utter evaluative statements.[148] Another, clearly related, reason is that closer attention paid to how the terms are actually *used* shows that we are not merely using evaluative judgments to ascribe properties. While the surface grammar of evaluative language might suggest some form of realism, then, the *function* of evaluative judgments has struck many philosophers as essentially *prescriptive* or *expressive*.[149] Value, it would seem, is a concept with mixed loyalties. We could even say that value has both non-cognitive and cognitive *aspects*, i.e. that it expresses both beliefs and non-belief-like mental states.[150] We face the choice between explaining away the non-cognitive aspect within a cognitive theory, or vice versa. Realists can say that we usually like good things; there are reasons to like them. So when we say about something that it is good, we usually simultaneously express our liking of them and recommend them to others. Non-cognitivists, on the other hand, can say that our attitudes have objects, and that we tend to associate and project our attitudes onto those objects, and talk *as if* value were actually a property of that for which we express our appreciation.[151] If no such reduction seems plausible, we might attempt a *hybrid*-theory, by allowing the domain to be split up in two or more components. Disambiguation offers a neat method to both diagnose and settle philosophical controversies.

[148] Mackie (1977), Hare (1981).
[149] Classic proponent/statements of this view is Ayer (2001), Stevenson (1937), Hare (1981).
[150] Hare (1981), Smith (1994).
[151] Blackburn's "Quasi-realism" (1993) is a theory of this sort.

In what follows, I'll presuppose merely that the possibility of 'value' being the name of a property is not ruled out. This is the minimum condition for naturalism. But neither is it ruled out that 'value' might perform multiple duties. Even descriptivist naturalists must admit that evaluative terms are frequently used to recommend, or to express some attitude or other.[152]

2.1.5 Disagreement

One of the most interesting features of value discourse is the existence of widespread and quite fundamental *disagreements* concerning most things evaluative. Some regard first-order matters, i.e. what things are good. Others are of a second-order nature, about those first-order predications and concern the meta-ethical issues mentioned above. The existence of disagreements in first-order ethics can be used as an argument in meta-ethics.[153] Seemingly irresolvable disagreements might be held to demonstrate something about the domain. A meta-ethical theory, in turn, can be used to settle, or at least diagnose, disagreements in first order ethics.

It is to disagreements in *meta-ethics* we must now turn. Can the extensive and seemingly irresolvable disagreement in *this* domain be used as an argument as well? The fact that most central meta-ethical statements are disputable suggests the following: A meta-ethical theory must take *some* stance or other on the issues on which meta-ethicists disagree, but no particular stance is mandatory to qualify as a meta-ethical theory.

Some of the disagreements in meta-ethics are such that it is hard to construe ones opponents as simply *mistaken.*[154] The problem is that 'value' as accounted

[152] Railton (1989) points out that the wise cognitivist allows moral language to play some prescriptive function. See also Putnam (1981).
[153] Non-cognitivists and relativists typically appeal to first-order disagreement, noting that realism/objectivism/cognitivism is hard to combine with the persistence of such disagreement. (Brandt 1998).
[154] See Smith (1994) and Darwall, Railton and Gibbard (1992) who points to this exact problem of fundamental disagreement.

for on one theory might be incompatible, not merely, trivially, with some conflicting account, but with what that account takes to be the pre-theoretical subject matter to be accounted for. The existence of rational disagreement shows that few if any beliefs in this domain deserve the status of being "self-evident", which, in turn, undermines their ability to serve as *foundation* for a theory. If we cannot find a universally acceptable pre-theoretical approximation of the subject matter from which we can attempt to reach theoretical agreement, we might have to change tactics. At some point, explaining the appearance of proper disagreement *away* becomes more plausible than respecting it.[155]

The fact that we construe disagreements in ethics and meta-ethics as *proper* disagreement suggests that there is, or that we *believe* that there is, a set of common beliefs about the subject matter. It might yet turn out that the appearance of a common subject matter is illusory, and that the subject matters of the disagreeing parties merely significantly overlap. Further, the agreement required for proper disagreement might not be sufficient to *settle* that disagreement. While it's desirable that an ethical theory solves practical problems, the truth might not answer to that desire.

While it is not *inconsistent* to treat a controversial value-relevant feature as a conceptual fact, it's unwarranted *when we are doing meta-ethics*. *Given* a certain meta-ethical view, any conflicting meta-ethical claim would be trivially inconsistent with the view assumed. In order to argue for any such a view, therefore, we need to take a step back and treat things that we might believe to be obvious as up for discussion. The fact that I, as a proponent of a certain meta-ethical view, claim that others are wrong about value does not mean that I think they are not doing meta-ethics.

[155] Brandt (1985) argued that moral philosophers should demonstrate that the pattern of concepts they propose has advantages for moral discourse, such as clarity, rather than try to capture some common sense notion explicitly.

2.1.6 Where do we begin?

In ethical theory and in meta-ethics, we need some common ground to start from to ensure that we are not speaking past each other. This common ground might consist of false or unwarranted propositions, we just need it to serve as a tentative basis for theoretical inquiry. There are a number of things we believe to be true about value, and a theory of value is a theory that somehow latches on to *those*. A theory of value does not necessarily have to make all widely shared beliefs come out *true*, though. I'm saying two things here: 1) these beliefs are our starting point; these are the things a theory of value needs to make sense of. 2) A theory of value needs to *somehow* account for these beliefs. There is no agreement about *how* this must be done for such an account to be acceptable as a theory of value. A case in point, which we will treat in some detail below, is the question of how value relates to *motivation*.

The common conception of value consists of the beliefs we have about value: it consists in what we believe to be valuable, the inferences we are liable to make about value, and about people making value judgments. It consists of what we already believe, what we take ourselves to have good reason to believe, and thus are reluctant to give up on.[156] Arguably, few if any of these beliefs are unconditional. Or rather: it seems that people, philosophers very much included, vary in how rigid they are in their beliefs about these things, and thus in what they will accept as a theory of value at all.

In approaching the fundamental problems of value, we should start with anything that looks promising, or, even better: with *everything* that does. The start, and to some extent, the end product of the theory presented in here is the statement that a number of features essential to our conception of value are enlightened by features of pleasure and hedonic processes. Our beliefs and intuitions about value can be traced back to pleasure. This is not merely a

[156] Smith (1994), Jackson/Pettit (1995), Lewis (1989), Railton (1989), among others.

matter of appealing to *substantive* intuitions, mind, but to *all* our intuitions about value.

The role of Intuitions

Classical intuitionism appeals mainly to our responses to real and imagined cases, and asks of us to categorise them as good or bad, right or wrong. The intuitions appealed to in *meta*-ethics are part of a much broader set of beliefs and belief-like states.[157] Clearly, the type of direct intuitive responses we have to events taking place before our eyes is quite different from the "intuitive sense" in which we might favour a property-like semantics over an expressive framework to account for the meaning of value statements. Only the former "have no further justification", and thus constitute the proper domain of intuitionism in its classical guise.

To say something counter-intuitive is always a cost for a theory. But how much so depends on, first, what else the theory can explain, and second: what role intuitions play in that theory. If intuitions are part and parcel of our epistemology for the domain, the cost of contradicting them is considerable. If, on the other hand, the theory postulates that substantive intuitions are likely to go astray and can provide a plausible account of how that might work, and the theory is still justified, counter-intuitiveness is less of a cost.

2.1.7 Analysis, explanation and justification

Is a theory of value supposed to *explain* anything? Whereas conceptual analysis has been the dominant strategy in philosophy in general and in meta-ethics in particular for the last hundred years or so, *explanation* seems to be just as important a theoretical notion.[158] Theories are frequently evaluated on account of what they can and cannot explain. I believe the case for meta-ethical

[157] Wide and narrow RE, see Daniels (1979), Tersman (1993).

[158] But see Harman (1977), Brandt (1985, 1998) Flanagan (1998), Copp (1990 (who does not think that confirmation theory, as he calls it, provides *justification* of moral standards), Railton (1998), Sturgeon (1985) More recently Joyce (2006), Stich and Doris (2006). These accounts have mostly focused on morality, rather than value.

naturalism depends on the success of the explanations it can provide. Or rather: the best case for naturalism is one that engages with explanation.

Explanation and justification

One sense in which the notion of explanation is important for our purposes is as contrasted with *justification*. Statements, mental states and beliefs can be treated within a theory as something to be explained, as well as something to be justified. The main purpose of ethical theory has often been presumed to be to find a standard of *correctness* for moral and other evaluative statements. *Explanation* of such states and statements, on the other hand, is the business of moral *psychology*, and need not entail the truth of the statement explained. Indeed, we often use explanations of evaluative beliefs as an excuse for failure of justification.[159]

The relation between explanation and justification of beliefs is not obvious, and nowhere is it less obvious than in the moral/evaluative case. In general, beliefs can be justified by their explanation being of the right sort, and undermined by them being of the *wrong* sort. My beliefs about the external world are, I believe, mostly justified by explanations connecting facts in the world to my beliefs about them. But value-theory/ethics, it has been argued, is an *autonomous* domain, disconnected from scientific explanations.[160] On this view, while our judgments and behaviours might very well be open for scientific explanation, this is irrelevant to ethical theory. "If you're in the explanation business, reasons look like a distraction; if you're in the reason business, explanations look like a distraction", as Kwame Appiah recently quipped.[161]

One possible reason for not engaging with explanation in ethical theory is that normative beliefs are not about how things are, but about how they *should* be, and this is what we need to account for. Causal explanations are beside the

[159] However: according to some epistemologies, and for some domains of belief, a causal connection between evidence and belief is *required* for justification.
[160] See Jackson (1974).
[161] Appiah (2008).

point, because no causal relation holds between what *should* be and our beliefs about it. Naturally, causal relations might hold between how things *are* and our beliefs, so when things are as they should be, such a relation holds, but this is not a further causal fact. So the argument goes. This makes the argument trivial, however: if you believe that evaluative facts make no causal difference, *of course* they won't appear in causal explanations. But we have not arrived at such a point yet, and as we are now looking for support for a theory of value, we need to consider seriously whether value can hold its own in proper explanations.

A more serious allegation is that *even if* there is a reliable causal connection between evaluative facts and evaluative beliefs, this is not what value theory is about. It is about *justification*, and the special nature of the domain is such that no causal explanation amounts to a justification. Justifying a belief involves citing reasons for holding that belief, as a rational response to the available evidence. Explaining why the belief is held, on the other hand, is a psychological project that entails finding out the *causal* processes involved in bringing the belief about. The causes involved in that explanation might be related to the reasons given for the belief, but then again, they might not.

If the goodness of a thing or a state of affairs plays no role in the production of the belief about that value, we might very well wonder whether value has a place in our world at all. If we don't need evaluative facts to explain anything, not even our evaluative beliefs, we don't need evaluative facts, period, and some other form of account should replace it.[162]

We need to remind ourselves that we are doing *meta*-ethics, not ethics here. We are not concerned with justifying first order evaluative statements, but with second order statements concerning the nature of value and evaluation. We are not, yet, saying - as naturalists are supposed to say, according to Nicholas

[162] See Harman's "The nature of morality" (1977).

Sturgeon - that morality and science are on a par: we are exploring the possibility that *meta*-ethics and science are on a par, a very different proposition altogether.[163]

The explanation of evaluative beliefs I will offer does not entail a justification *of those beliefs*. It supports a meta-ethical view about the nature of value that is in fact incompatible with a large set of common evaluative beliefs. The explanation offered will establish a referent for evaluative terms, but not by making our most common beliefs about value true, nor by capturing our ordinary justificatory efforts. On the contrary, the explanation involved will demonstrate that justifications often tend to lead us away from the evaluative facts. In this sense, explanation is indeed at odds with justification. The explanations appealed to undermine some justifications, but it does not undermine evaluative beliefs in general.

Explanation and conceptual analysis

A further reason to engage with explanatory matters, largely overlooked in the literature, is the following: Regardless of what analysis you favour, there are value relevant facts that should *not* be assigned conceptual status, and still need to be accounted for. Most of our substantial intuitions about what's good seem to be of this sort. We are pretty sure that some things are valuable, but we don't usually treat it as a *conceptual* fact that this is so. We can, of course, require that a theory of value delivers approximately the right set of substantive goods, but it is not a *conceptual* requirement. Contradictions in terms are not the only theoretical shortcomings, and consequently, conceptual matters are not all that matters.

The same point applies to the relation between value and *motivation*. The disagreement between internalists and externalists concerning how evaluative

[163] Sturgeon (2005). If meta-ethical naturalism is true, however, it *follows* that moral statements are factual, and thus "on a par" with scientific statements. But that is not yet the proposition under assessment.

beliefs relate to motivation suggests that no *particular* relation is conceptual. *Some* relation is probably required for a notion of value to be recognisable as such, but what relation this is, and what's its status should be kept open. Alternatively, it's a conceptual truth *that* value relates to motivation (and that positive/negative value relates to positive/negative motivation) but not what *particular* relation that is: *that* may be an empirical question, to be settled by any theory of value aiming at being complete.[164]

Conceptual analyses can be *enlightening*, but *qua* analyses, they do not *explain* anything. Analysis *relates* to explanations insofar as what an analysis does *not* say about its subject matter, but what we nevertheless believe to be true about it, is something we need to explain. The more inclusive the analysis is, the less need do we have for explanation. Moreover, if an analysis covers a certain notion, that notion can no longer be informatively invoked in explanations of the feature in question. You cannot invoke the bachelorhood of a man to explain why he is unmarried (whereas an independently established fear of commitment just might do). Similarly, a theory that makes motivation a part of proper evaluative judgments cannot explain why, or how, such a judgment motivates: it wouldn't qualify as an evaluative judgment if it didn't. On the other hand: A theory that does *not* link motivation to value conceptually *needs* to explain why (ascriptions of) evaluative properties often motivate.

To doubt a conceptual analysis is, arguably, to refute it as an analysis, provided that the doubter is recognised as a competent user of the term. Explanations, on the other hand, do not need to be obvious in order to be successful: their application is not contingent on their accessibility to anyone competent with the term. Explanations have the power to *persuade* and *convince* people, due to the fact that they are allowed to bring something *new* to the discussion. Analyses are not supposed to do anything but state what ought to be obvious

[164] Frankena (1958): The question is whether motivation is somehow to be "built into" judgments of moral obligation, not whether it is to be taken care of in some way of other". This disagreement cuts across most other issues in meta-ethics, and might be more basic than those.

already, or at least on reflection *ex post*, and the argument from disagreement aims to show how little can be accomplished by following that route.[165] With less than catching them being outright inconsistent, we are unlikely to persuade opponents to accept our rival analysis. But the explanations a theory provides can do precisely this. We might even be able to persuade people to accept a controversial analysis on basis of the explanations it *affords*.

A theory of value should be able to provide an explanation of that which is troubling about value, which is more or less everything about it. This is not limited to giving an account that fits with pre-theoretical intentions and inclinations. The fact that we are competent users of value terms does not mean that such an account is easily accessible to us. If we approach value from the perspective of explanation, we can treat value as more akin to a *scientific* problem, which opens up possibilities to get around apparent disagreements.

2.1.8 A note on the Epistemology of Value

The questions we are dealing with can be framed in *epistemological* terms. How do we acquire beliefs about value? How do we *learn* to apply value terms? Is our knowledge about value *a priori* or *a posteriori*? Do we know about value *innately*, infer it from other sorts of information, or acquire this knowledge by somehow interacting with the world? Is it analogous to knowledge of mathematics, or to empirical knowledge? The epistemology of evaluative belief also addresses the place for *causes* in the production of knowledge about value. This is, regrettably, not the place for a prolonged discussion of the deeper issues in epistemology, but some things should be said about it as applied to the evaluative domain.

Sui generis and practical knowledge

An influential argument has it that value is a *sui generis* property, a category of its own. One premise in this argument is that our knowledge of value is *self-*

[165] An analysis can be derived by features common use not obvious to the user, but should be on reflection. See Smith (1994), and below (chapter 2.2).

evident, and thus not based on/inferred from our knowledge of other things.[166] This domain might have a unique character, something to be *invented* or *discovered*, rather than borrowed from some other domain of knowledge. We could then no longer draw upon any analogy but analogies are usually flawed methods for understanding anyway. If 'value' were thus a *simple* and *fundamental* notion, this would explain why it's so hard to be specific about it, and to provide evidence for its presence. Moore, famous for defending this view, did in fact say more specific things about goodness: The simplicity of the notion, he argued, merely implies that what else is true about goodness does not form a *part* of it. 'Good' is not something we understand by grasping *other* notions: it is a primary notion that we know about *non-inferentially*. But if this is all there is to it, we are stranded when facing disagreements. Granted, self-evidence does not mean *obviousness*, but whatever there is left to it to mean needs to enable dealing with disagreements.

Our knowledge about value could be *practical* knowledge, to be understood in analogy with *skills.*[167] If this were the case, whatever theoretical framework we settle for would be something of an afterthought. *Practicality* could be what distinguishes the subject, and *any* theory concerned with practicality in a suitable manner would thereby qualify as a theory of value/morality. There is, I believe, some truth in this, and indeed, the idea that this field is concerned with *knowing what to do* has some appeal, even though some qualification as to the reasons for action involved seems necessary. Seeing how practicality has often been held to be the failing part in any realist theory, this might be a fruitful strategy for incorporating this element.

[166] Judgments about the value of particular things are commonly not supposed to be self-evident, since their possession of good-making characteristics might not be obvious.
[167] Like grammar. Rawls used this analogy, and it was developed in Hauser (2007). Frans de Waal (1996) points out that morality, like language, is to complex to be learned by trial and error, to variable to be entirely genetically programmed. Rather, it is a mixture of both.

Causal interaction

Do our beliefs about value depend on causal interaction with the world? And, seeing how the causal story about a belief can confirm it, undermine it, or leave it perfectly intact, which is it in this case? Some species of belief are such that a casual relation to the object of belief is *required* in order for it to qualify as knowledge. Furthermore, the causal relation must be of the right kind: Beliefs does not qualify as knowledge (or as *true*) merely because there is a reliable connection between some fact and the formation of the belief: the fact causally responsible must be the fact believed in.

Some beliefs about value are probably innate. While our ability to categorize things as good or bad, and to communicate about it, might improve with training, this involves improving on a notion we have some grip on already. This doesn't mean that these beliefs are "a priori", however: it means that the relevant causes are not all *external* to the judger. I'll argue for an *empirically responsible* theory of value, and this will involve turning to the *cognitive* and *affective* sciences. The reason why is not merely because cognitive processes like belief formation and weighing of evidence are on the *receiving* end of the causal chain[168] but because such processes are at the *transmitting* end as well: The causes of evaluative beliefs come at least partly from within.

Contingent starting point

What kind of knowledge we have about value depends on what question we are asking. In a theoretical inquiry, we can treat more or less anything we believe as established, fixed, and others as put in to question. According to one influential theory, more or less anything we take ourselves to know might turn out to be false, definitions included, and almost anything can turn out to be necessarily true, given that we treat the conditions as fixed.[169] It *looks* like conceptual facts are fixed, and empirical ones are contingent, but a closer look shows that this can change in the light of evidence or new theoretical considerations. We can

[168] This was the argument behind naturalized epistemology, (Quine 1974).
[169] Quine (1978).

treat any belief as a "fundament" of our theory, and there is no further justification for favouring beyond theoretical virtues like simplicity or conservation of commonly held beliefs. The importance of this observation is meta-theoretical: When we ask first order questions about what things are good, we usually take some notion of value for granted. When we ask questions about the *notion* of value, on the other hand, we typically do not. The argument from disagreement shows that we should be prepared to treat most beliefs about value as not yet settled. But this shouldn't keep us from settling them tentatively, to suspend disbelief until we have seen what the resulting theory can accomplish.

There are a number of beliefs about value that the theory needs to account for. To say that this cluster of beliefs should *somehow* be accounted for is to treat their epistemological status as open until further notice. Some beliefs might be given precedence, so that we think of them as *more important to keep true*, while other beliefs can be discarded given sufficient reason. This returns us to the question whether there is anything *essential* to value, i.e. whether there are any central beliefs that a theory of value *must* make true if it is to be a theory of value at all.

2.1.9 Value-theory naturalized

I'm concerned with *naturalizing* value theory. As with naturalized *epistemology*, the relationship to which should be apparent, this involves a programmatic replacement of a "purely philosophical" approach with an inquiry sensitive to empirical findings. Primarily, this involves taking stock of the processes involved in forming beliefs, desires and motivations. Hence, we are interested in the *psychology* of evaluation. Naturalizing value theory is not a direct move to naturalism in the metaphysical sense: investigation into the psychology of evaluation could just as well lead to the abandonment of value realism. But the investigation presented here *does* result in a vindication of naturalism: naturalizing value theory is compatible with, indeed supportive of, reductive naturalism. Naturalized value theory is concerned with explaining what's

troubling about value, and with doing so without leaving the natural domain.[170] It could also involve the claim that there are no further truths of the matter: naturalized value theory is all the value theory we need.

The case for naturalism relies on value-theory being roughly analogous to scientific inquiry. The argument for naturalism is often made with reference to the identification of water and H_2O. This identification was not established by conceptual analysis, but by empirical research: the identity is something we *found out*. But this identification, surely, presupposes *something* about the concept in question; that property identity follows from functional equivalence, perhaps. The same is not obviously true for value terms, which means that we cannot just help ourselves to such a claim.

In what follows, I'll be concerned with answering this, and other challenges. But it should be realised that this is mainly accomplished by putting the matter of justification to one side, for now. Justification need not be given, indeed might not be possible, in advance. Our main concern is not whether what we come up with could seamlessly replace our current talk about value, but whether everything important, or enough of it, can be accounted for. It's possible that naturalism can only succeed on its own terms. If the challenge is that something above and beyond the listed explananda, functionally arranged and defined, is lacking, the naturalist can reasonably doubt whether there is such a thing, or that value theory needs to account for it.

In *Philosophical Naturalism*, David Papineau points out that while naturalists treat philosophical problems as continuous with the problems in natural science, they are different in kind. They are characterized by

> a special kind of difficulty which means that they cannot be solved, as scientific problems normally are, simply by the uncovering of further empirical evidence. Rather they require some conceptual unravelling, a careful

[170] Slote (1992). Flanagan (1998), Appiah (2008).

> unpicking of implicit ideas, often culminating in the rejection of assumptions we didn't realize we had.

While not being part of everyday scientific practice, such questions are part of scientific inquiry in the formative stage, and in its most inclusive sense. In the beginning of the construction of scientific theories, Papinau writes, "the task of philosophers is to bring coherence and order to the total set of assumptions we use to explain the empirical world". A doctrinaire philosophical naturalist would perhaps say that this is *all* philosophical theorizing should be.

In his book "Thinking how to live", Alan Gibbard pointed out that his account might in fact not be strictly true about our actual normative concepts. Nevertheless, he argued, the concepts he described seems quite useful, and are strikingly similar to our normative concepts, so perhaps our normative concepts are actually like that?[171] Would anything be lost if we *replaced* our existing framework with concepts of this kind? Would something, clarity for instance, be gained? In the absence of something better, we have reasons to accept a theory that, while not mirroring precisely the concept as normally used, is at least clear about what it is doing. But note that two "incompatible" accounts of value can play this game.

Reductionism and eliminativism

The question should be raised whether naturalizing value in the way proposed does not in fact *eliminate* value-theory, by making it a part of psychology.[172] Reductionist versions of naturalism should be sensitive to this challenge. Value, according to the view developed here, is a psychological property, and the things explained are primarily psychological events. The ambition of the theory is to bring about a "tolerable revision" of value theory. [173] It relies on the truth of certain psychological theories of motivation, learning, and, to some extent, concept formation. We do not need *sui generis* non-natural value properties in

[171] Brandt makes the same point (1985) .
[172] Slote (1992) is eliminativist, rather than reductionist.
[173] Railton (1989), Also, Brandt (1979).

order to explain "value" facts. Nevertheless, there is a property that does a significant amount of work in most of the relevant explanations: pleasantness. A psychological account is objectionably eliminative only if it involves no reference to essentially evaluative properties. And the claim is that pleasure *is* evaluative in nature.

2.1.10 Reductionist Hedonism

My ambition is to define and defend a version of naturalistic hedonism about value. The hedonistic view proposed is fairly ambitious; the objective is to provide a reductive account of value. Some controversial claims must be accepted for hedonism to be plausible, and I will make a case for accepting those claims, but won't pretend to make a conclusive one. To echo Mill, I can offer no "strict proof" for my thesis, but I will present some considerations in favour of the theory.[174] The claim that pleasure merely *has* value, or even that it uniquely has it, strikes me as too weak. The facts about pleasure that inform the hedonistic set of intuitions afford a much more ambitious claim. In terms recently employed by Crisp: the theory I propose is not merely hedonistic in the *enumerative* sense, it is hedonism in an *explanatory* sense. Pleasantness is what *makes things good*. This is not merely a supervenience claim; it's an *identity claim*. This is characteristic for reductive naturalism: if goodness is a natural property, having that property is what "makes" that which has it good. I'll argue that pleasure is the common element with reference to which we can make sense of evaluative language and practice.

I will offer an *explanation* of key evaluative features in which pleasures, or 'hedonic processes' play a crucial part. Hedonic processes, I will point out, are absolutely central to our evaluations, motivations and behavioural tendencies, and this is what justifies the hedonistic approach to value. The argument is thus strikingly similar to the classical, and classically rebutted, move from

[174] Mill *Utilitarianism* (1993).

psychological hedonism to value hedonism.[175] The work cut out for the hedonist is to argue, partly for pleasure playing this central role and partly to argue that this is sufficient to establish an identity claim. Our beliefs about value can ultimately be said to *track* pleasure. Now, to say that our beliefs about value can be shown to stem from hedonic processes is not necessarily to say that pleasure and value is the very same property. To track down the causes of a belief can, as mentioned above, undermine, rather than vindicate, those beliefs.

Given a certain naturalist framework a hedonist version of that theory can be substantiated. An argument *for* that naturalistic framework needs to appeal to further considerations. There are some quite general benefits of naturalism: such as being included in a highly successful and potentially unified theory of knowledge.[176] But the argument for such a framework need not be prior to, or independent of, what it *turns out*.

Hedonism is a contested theory, often criticised for being too simplistic; unable to assign value to things that clearly have it. The hedonist must be able to explain such beliefs away. The theory will quite generally have to explain away evaluative principles that are, in Hare's phrase, "strongly internalised"[177]: i.e. principles that, while having an explanation not supporting their truth, are hard too be conceived of as false. This process of internalisation might, admittedly, influence the content of the concept so that it's meaning is not a mere name of the property from which it stems. But a sufficiently powerful explanation in terms of that property might still warrant the naturalistic identity claim. A property can cause a concept whose content might then transcend its cause. *But* if there is no clear distinctive meaning of that developed concept, or if the meaning is a contested one, this property might yet be what provides the characteristic, "core evaluative" meaning to that concept. If we can show how disagreement about the meanings of evaluative concepts results from a

[175] Moore (1993), Bradley (1962).
[176] This is Boyd's point in "How to be a moral realist" (1988).
[177] Hare (1981).

psychological process, the common core is to be found *before* the divergence. Pleasure could be conceived of as *proto-value*; i.e. as the *most basic evaluative phenomenon*. This beginning might be the last point at which we can expect any commonality between competing theories: it is the source of the appearance, valid or not, of a common subject matter.

The argument is that pleasure is causally and constitutively responsible for our beliefs about value. But I will also claim that it is also evaluative *as an* experience. This was one of the reasons for choosing the proposed analysis of pleasure. It has direct intuitive support *and* causally, reliably, regulates our beliefs about value. Our *irreducible* grip on value consists in the *experience* of value. Pleasure, I'll argue, *is* that experience, it is not (or not just) the object of the experience. The second approach defended is to say that "value" is the thing *answering to, or causally responsible for*, our beliefs about value. There are a number of things we believe to be true about value, and value, accordingly, is the thing that those beliefs tend to be true about *and/or what causes those beliefs*. The best available natural property playing such a role is pleasure. Such a theory would seem to be *a priori* naturalistic, and merely *happen* to pick out pleasure. However, the fact that pleasure is fundamentally evaluative, and (partly) responsible for those beliefs clustering the way they do, suggests that the hedonistic part is actually primary. This is of importance in dealing with some objections towards the type of naturalism proposed.

Value can be conceived of as a natural property. Whether or not 'value' *is* a natural kind term it can be *treated* as such. If we do so, we can offer a strong case for hedonism. So *what if* we treat value in a way posed to support hedonism? Do we get a theory that exhausts the domain? Is the resulting theory helpful, practical, enlightening, in any way? I believe it is, and that no matter the status of the theory, this is a theory worth developing.

The limits of the theory

The hedonistic position argued for is a theory of *value*, understood as synonymous to *goodness*. It is not, or not directly, a theory about *rightness,* or *rationality*. I'm convinced that we have good reasons to accept the hedonistic meta-ethical position as presented, but not whether these are conclusive *reasons to value*, in the sense "appreciate", pleasure, nor do I rule out reasons to value other things for their own sake. The theory defended does not yield moral conclusions. In accepting it, we have to give up on the claim that value-theory caters in a decisive manner to other fields of normativity. Normative answers, if there are any to be found, have to be found elsewhere. Nothing I say implies that the right action, or the rational action, is the pleasure-maximising action. Indeed, as we'll see, there might not even be an unambiguous way of evaluating outcomes on this theory. There most likely is a connection between what's good and what's right, it might even be of a conceptual nature, but I offer no theory about it.[178] We would be *right* to accept hedonism, not righteous. This view will be expanded in chapter 2.5.

Plan

Over the next few chapters, I will describe and argue for *naturalism*: first by pointing to the nature and benefits of naturalism in general (chapter 2:2), and then by comparing and drawing the lesson from two contemporary naturalist approaches to meta-ethics (2:3). In chapter (2:4), I will make a case for empirically informed value theory. The appeal to relevant sciences, and to the claim that some sciences are, is due to two points. First: we need all the resources we got in order to account for philosophically troubling issues. Since a lot of the concepts invoked in meta-ethics have an empirical, psychological aspect, we should let our theory engage with the cognitive and affective sciences. The second reason is that if we are naturalists of an a posteriori inclination, *of course* natural science will have a role to play. Hedonists in particular would benefit from an empirical approach to value theory. The last

[178] Williams argues that moral philosophy cannot deliver the thing we expected from it, i.e. a guide for ethical reasoning. Williams (2006).

chapter (2:5) develops the version of hedonism the rest of the book paves the way for. In short, I carve out a value theoretical position, based on the centrality of pleasure in the explanations of value-relevant facts.

2.2 Meta-ethical Naturalism

2.2.1 The nature of value

> In trying to improve the unity and economy of our total theory by providing resources that will afford analyses… I am trying to accomplish two things that somewhat conflict. I am trying to *improve* that theory, that is to change it. But I am trying to improve *that* theory, that is to leave it recognisably the same theory we had before. – David Lewis, On the plurality of worlds

Naturalism in meta-ethics, broadly conceived, is the view that an account of ethics, or value, does not need to leave the natural realm.[179] Naturalism in this wide sense would include most versions of non-cognitivism, error theory and relativism. Indeed, it can be argued that those theories are based on a commitment to general philosophical naturalism: the natural world leaves no place for value properties understood realistically, so if the evaluative domain shall be accounted for, we need to back off from that particular claim.

In this chapter I will argue for and about naturalism in a narrower, but historically more significant, sense of the term; the view that value is a natural property. This view is often held in conjunction with naturalism about properties in general, but the meta-ethical naturalist is not required to believe that all properties are natural.[180] Naturalism about value is logically independent of general philosophical naturalism. But it is not *theoretically* independent: Under suitable circumstances, as we shall see, philosophical naturalism *supports* meta-ethical naturalism, and under unsuitable

[179] Lenman (2006).
[180] See Copp (2003.)

circumstances, it undermines it. In addition, the case for meta-ethical naturalism is undermined if philosophical naturalism is false, or incomplete.

Naturalism about value is also logically independent of the existence of *other* non-natural facts and properties, but again, there is a theoretical relation[181]: The fact that naturalism provides objective value facts without adding to the ontology should be considered a reason to accept it, but this reason weakens somewhat if there are precedents of non-natural properties. Value properties are without doubt philosophically *elusive* properties, and the difficulty in accounting for them as natural properties serve to fuel the anti-naturalist's case. If we do accept other non-natural properties, then, we need to argue why value should be camped with the natural rather than the non-natural properties.

The view I defend is a *metaphysical* thesis: it claims that value is a particular natural property. It is not, primarily, a semantic thesis. While being a kind of reductionist (to be described later on), I do not offer any reductive naturalistic *definitions* of value terms. Or, rather; I do not claim that any such definition is analytically true. Naturalism is not true solely in virtue of the meaning of value terms.[182] Insofar as there is a "common-sense meaning" of 'value', it is such that it *allows* treating value as a natural property. Neither naturalism nor any of its main competitors is true or false in virtue of the common meaning of evaluative terms: none of them is strictly abusing language.

I believe that a successful metaphysical theory of value, completed by a plausible epistemological story, can justify adopting a naturalistic form of semantics for value terms. This would involve offering a more specific meaning to the term than it ordinarily possesses, and thus, to some extent to *change the subject*. Such a theory would have to earn its place as an account of value in naturalistic terms.

---

[181] Cases in point being *mathematical* facts and properties, see Sayre-McCord (1988) and Copp (2003).
[182] This puts me apart from naturalists who believe that *general* naturalism is analytically true, but that no *particular* identity holds analytically. Jackson (1998), See Smith (2004).

2.2.2 Natural properties

To understand what treating value as a natural property entails, we need some idea of what natural properties are. This has proven notoriously difficult, but here's an attempt at a working definition: Natural properties are characterized by what they *do*, i.e. by the causal difference their implementation makes - or the difference they *potentially* make, if they are understood as dispositional.[183] Natural properties can be identified via their *function*. Being of a certain weight entails exerting a certain pressure on the surface on which you rest. To identify what weight is, then, you need to identify what plays this role, i.e. find out what the conditions are for exerting pressure. The most common example in the literature about natural properties is the property of being *water*.[184] Water is identified as that which plays the "water-role", which, it turns out, is the chemical compound H_2O. Natural properties can also be defined as *functional* properties, i.e. not identified as *what* plays the role, but as the role itself. Water, on such a reading, would not be the property *that* plays the water role, but the property *to* play the water role. The difference surface when we consider possible scenarios where something *other* than H_2O plays the water role.[185] If performing a function is *essential* to a concept/property this latter mode of identification is preferable: we would want the function to carry over to counter-factual scenarios. If the function is merely how you *identify* the property, on the other hand, and considered contingent or accidental to the property or concept in question, the former reading is more fitting. *Names* are usually held to be of this sort, whereas terms such as 'edible' are of the other, functional, kind.[186] We will return to this distinction in greater detail in the next chapter.

---

[183] Rubin (2008) on beliefs about natural properties being sensitive to a posteriori investigations, Copp (2000) on "empirical" properties, Moore's notion wasn't very specific, but seems to be in line with these suggestions.

[184] Water is not merely an example of a natural property, but of a natural *kind*, or *stuff*, for which special rules applies. See Putnam (1975), Boyd (1988) Copp (2000).

[185] Putnam (1975). For application to the moral case, see Sayre-McCord (1997) and Copp (2000).

[186] Kripke (1974). On this distinction between functional kinds and natural kinds, see Kim (1997) Sosa (1997), Copp. (2000).

If we intend to treat value as a natural property, then, what does it do? Is there a causal function value should perform? Admittedly, to say that value *itself*, rather than the good-making features of that which has it, does anything at all is to beg some crucial questions.[187] Non-naturalists typically claim that value itself has no causal function and that it merely *depends* on properties that do.[188] But naturalists should be adamant: if value makes no difference, what do we need it for? Even mathematical properties, whose irreducibility has some theoretical acceptability, earn their status due to their role in scientific explanations. What comparable benefits are there to value properties?[189]

We might ultimately have to beg the question, but, for now, we can afford to ask it hypothetically: *If* we treat value as a natural property, what function should it perform? Can the things we *need* value for be functionally specified? Whether or not performing such a function is essential to value, whatever its role in our theory turns out to be, is there one?

If natural properties have causal functions, one step towards identifying them is by looking for typical consequences. We have touched upon the possibility that values cause motivational states, but cannot make more substantial claims about that link yet. Whatever else might be true about value, one of its natural "consequences" seems to be the existence of evaluative mental states and statements, however those are to be understood. So, are there any kind of indication that the occurrence of these states and statements vary with empirically verifiable conditions? We could, in effect, treat occurrences of evaluative states and statements as *detecting* the presence of value. We could start out by looking for typical causes that might serve as *predictors* of these

[187] See Nagel (1986). The "good-making" relation means that goodness holds in *virtue of* other, often natural, properties that the good thing has, but that these properties are not *identical* to goodness. Hoping to circumvent controversies, this is more or less equivalent to the *supervenience* relation.
[188] Values still serve *some* function on those account: to provide normative reasons for pro-attitudes (see Nagel 1986). I'm here talking about *causal* functions, however. How reasons relate to attitudes *causally* is a complicated matter, not dealt with here.
[189] Even if value *properties* serve no causal function, value *statements* might, of course. But that is, presently, beside the point.

states. Using *colour* as our model, we could then identify what causes those states/statements and brand the typical cause, if there is one, 'value'. In general, if evaluative claims somehow *behave* as though they are sensitive to a posteriori findings, this would speak in favour of giving value the full natural property treatment. And even if this is *not* enough to secure the identification of value with that causal property, it would at least suggest that there is something going on worthy of a naturalistic, scientific, investigation.

2.2.3 The desirability of naturalism

Ontological parsimony is a philosophical virtue: if we can do without the addition of another realm of properties, we should. This principle favours ethical naturalism, insofar as naturalisms in other areas of knowledge form a cogent and lucid whole. Now, having said that, there are reasons to believe that we just *cannot* do without positing non-natural properties to account for value. One such reason is the fact that most people believe that some of our value judgments must be *true*, and if we are persuaded by any of the anti-naturalist arguments below, non-naturalism would seem to be our only option. There is, however, something odd about a property whose sole function is to act as the truth maker for a certain sort of judgment (an epistemic relation, not a causal one, mind). Granted, non-naturalists normally think values perform the further function of providing *reasons* for actions and attitudes, but this function is not supposed to be a causal one either, and thus not part of the best *causal* explanation of events. Positing non-natural properties should not be a theoretical first choice.

Naturalism is desirable not only for its ontological parsimony, but, as we mentioned, also for its *explanatory potential*. To say that naturalism is desirable is to say that it has some *purchasing power*: we should be more willing to change our beliefs to accommodate naturalism than we should to accommodate something like non-naturalistic realism. Whether the theory ultimately leaves

something essential unaccounted for will have to be decided in retrospect, once we've seen what a naturalist theory can do.

Value terms might refer to certain natural properties, and relating value predications to empirical conditions that reliably produce such predications would seem to be a perfectly respectable way of ascertaining such a reference relation, and a worthwhile project in its own right. But it should be recognised that all terms that somehow refer to the same particular property are not necessarily synonymous: they do not necessarily express the same concept.[190] There could very well be irreducibly normative concepts "for" natural properties. Gibbard argues that while the property water is the property H_2O, the *concept* water is not the *concept* "H_2O". [191] Similarly

> "…whereas the *concept* of being good is distinct from any naturalistic concept – from concepts fit for empirical science and its everyday counterparts – the *property* of being good is a natural property, a property for which we could have a naturalistic concept. (p323)[192]

Gibbard thinks that normative concepts ("value", "right", "ought") should be analysed in terms of "the thing to do", and he believe that some natural property *constitutes* "being the thing to do". Non-cognitivists are usually naturalists in the sense that they deny the existence of an *irreducible* realm of moral properties and, Gibbard aside, they tend to say that *strictly speaking* there are no moral properties. There are natural "moral" properties in the trivial sense that some natural properties have moral importance: this is just what the supervenience claim says. They could even acknowledge that moral terms *refer* to these natural properties. Similarly, non-naturalists who believe that value supervenes on a *general* class of natural properties can admit that value terms co-refer to the supervenience basis for the non-natural property 'value'. Non-naturalists and non-cognitivists are not required to deny that some interesting

[190] Gibbard (2003, 2006).
[191] Gibbard (2006).
[192] Gibbard offers this as a development of Moore's distinction between the properties that *makes* something good and the property "goodness". Gibbard's account has the benefit of not adding further properties, but only concepts.

regularity between evaluation and natural facts can be empirically established.[193] What makes those views different from naturalism in the sense developed here is that moral importance, or value, according to non-naturalism is not one of those natural properties. It's a simple enough distinction, but one over which much of the last 100 years or so of meta-ethics has been fought. The distinction between what *is* good, and what *makes* something good on the one side, and the property of *goodness* on the other makes sense, even to the naturalist, but a different story is offered about it on that account. Naturalism can, as shall become clear, take the distinction into consideration and still offer an account that eliminates its significance: Naturalists can understand the distinction without accepting the consequence that the distinction carries a difference.

One of the main problems for meta-ethical theories is that they tend to presuppose what they are supposed to prove. Naturalism is a prime example of this tendency: *If* a general naturalistic approach is acceptable, we can develop naturalistic theories about value that look quite promising. But whether we *should* apply such an approach, is not that easy to establish. Nor is it that easy to refute. Whether naturalism can be the answer to the questions of meta-ethics and value theory depends, ultimately on what we take those questions to involve and, as noted, there is no consensus here, not even on what would *settle* the matter. There is no consensus about what "parts" of the concept must be respected, and what parts we might allow to be explained *away*.

Naturalism can, I believe, provide plausible answers to most of the central questions in meta-ethics, and it can propose how the questions *not* answered can still be dispensed with. I do not pretend to prove whether this naturalistic approach is the right way to do value theory, but some considerations can be offered in its favour. There are certain *benefits* to naturalism, but to see them as benefits might require a favoured explanatory model, or a certain ontology, the arguments for which are unfortunately beyond the scope of this work. My

[193] I apologize for not defining these positions clearer than I do. Doing better would take us too far from the point.

point, however, is that such arguments are not needed in order to make a *conditional* claim: *if* we take value theory to offer up a certain set of problems, naturalism can provide a solution. Ultimately, a theory can succeed only in what it aims at doing.

2.2.4 Methodological naturalism

It's important to distinguish naturalism as a *method,* from naturalism as a substantive theory of value, i.e. as a theory with naturalistic *content.*[194] The theory I aim to develop is naturalistic in both senses, and treats them as mutually supportive, but the two are in principle separable. This "in principle" separability is fraught with complications, however: Methodological naturalism does not necessarily lead to a theory with a naturalistic content, indeed, it has been thought to *undermine* such theories, but it would be decidedly strange if value was a natural property, but not open to naturalistic investigation.[195] Admittedly, to *establish* that value is a natural property does not necessarily involve methodological naturalism: some naturalists are *analytical* naturalists, after all. But once you have established a property identity reduction, value could be naturalistically studied. That, it could be argued, is not the concern of meta-ethics: the job ended with the successful identification. As will become increasingly clear; I strongly disagree with that argument.

You can, as we said, be a meta-ethical expressivist and still be a methodological naturalist. And non-naturalist realism is arguably based on the observation that no natural property could do what 'value' would have to do.[196] This has commonly been construed as a *conceptual* argument, but could also be understood as an *empirical* observation, in which case it would be a case of

[194] See Railton (1989), Doris and Stich (2006), Nichols (2004), Joyce (2006).
[195] Harman (1977, 1986), Joyce (2006, 2007), see ch. 2.4. Perhaps the claim can be made that meta-ethics is about the *concept*, not the property, of value, and that the concept "value" is not equivalent to the naturalistic concept of that property.
[196] Gibbard (2003) Mackie (1977), see also Joyce (2006). As Toni Rønnow-Rasmussen has pointed out to me, the expressivists were usually concerned with moral *language,* and not with metaphysics, and might disagree with this recounting of events. Possibly, the turn to language was based on a general disappointment with metaphysics.

applying the naturalistic method with a negative result. The matter is perhaps debatable whether it is an empirical or conceptual fact what natural properties can and cannot do.

Simplifying a little, methodological naturalism is the view that observation is a valid method for investigating and verifying propositions about value.[197] This can be applied to both first and second order propositions. Ethics, according to methodological naturalism, should not be construed as an *autonomous* domain of inquiry.[198] Methodological naturalism is the conviction that ethical and/or meta-ethical issues can be approached using broadly *a posteriori* methods, continuous with those employed in natural science. Empirical observations might enlighten what things actually are important for our notion of the good. Value theory might still be conceived of as a special domain, having a certain distinctive focus, and thus it need not be *reduced* to some *other* natural science, even if the property of value turns out to be shared with other sciences. Naturalizing value theory does not necessarily involve reducing value to be defined in terms of some other science: the concern is rather with widening the field of research, and kinds of questions open for scientific inquiry.

Self-reports as empirical observations

There are things we *believe* to be important for evaluative judgments, which we can investigate simply by asking what seems to us critical to the notion. What things, actions and events do we consider to be good, and why? What do we believe would make us change our minds, and do we believe that a change of heart would be warranted under those circumstances? The method to find *those* things out would seems to be accessible from the proverbial armchair; possibly the method is extendable by doing a survey. Whether to treat such a procedure as part of the "naturalistic" method is perhaps debatable, but most natural science latch onto statements about observations at *some* point, observations

[197] Boyd (1988, 2003) answers the question "what should play the role of observation in ethics" quite simply: Observation. Our evaluative judgments are based in experience as much as other beliefs are.
[198] See Jackson (1974), Sayre-McCord (1988).

about ones self very much included, so there is no particular reason to exclude it from the naturalistic approach. The beliefs we are dealing with are frequently called *intuitions*, commonly treated as essential in philosophical justification.[199] To treat these beliefs and intuitions in a naturalist manner is to treat them as fallible parts of the scientific data to be accounted for within the theory, as something to be explained, and not as part of the effort to justify beliefs from indubitable bases. [200]

One important benefit of approaching value with a broadly defined naturalistic method is that we can give an important theoretical role to factors that *turn out* to be important for evaluation, but that we typically don't foresee or intuit. I.e. there are factors *not* typically part of our offered justifications, and reported intentions. If we investigate what *actually* (i.e. *causally*) determines our evaluations, i.e. for evidence not already accessible to introspection, or by appeal to "common sense" or to semantics, we might be able to work our way past the current stalemate in meta-ethics. This involves engaging with, or at least paying attention to, psychological and sociological research. Importantly: listening to such research shows that what's important for evaluation is not always what we might think.

Now, can a case be made that such observations are of importance for meta-ethics? A number of accounts of evaluation, posed in psychological and evolutionary terms, have argued that identifying the causes of evaluative judgments would have a *deflationary* effect on ethics.[201] On the contrary, I'll offer some support that this process would result in the vindication of at least some evaluative statements.[202] Indeed, seeing how one of our strongest beliefs about value is that some such statements are *true*, deflationary accounts have a

[199] On intuitions as evidence, see Goldman (2007). Their main role is in *foundationalism*, which treat intuitions as the only possibly *end-point* of justification (Ross (2002) Stratton-Lake (ed.) (2002), but intuitions play a role as the *start* for Reflective Equilibrium theories as well. Rawls (1971), Tersman (1993).
[200] The for-runner for this strategy is arguably Richard Brandt (1998).
[201] Joyce, (2004) Harman (1977), Mackie (1977)
[202] In line with Railton (1989), Boyd (1988) and Katz (2008).

considerable theoretical cost.[203] Non-cognitivism and error-theory are certainly possible positions (actually, I believe there is some truth in both): The truth-desideratum is defeatable, as are *most* beliefs. But naturalism, if it can be made to work, is, as we say, desirable. Voices have been raised that philosophy can no longer afford to ignore the results of scientific findings, especially within psychology, and that therefore, if not *entirely* naturalist (we're not empiricists by default, after all), any philosophical effort needs to incorporate, connect to or, minimally, be consistent with natural facts about our subject matter, and about the process by which we reach our conclusions. Epistemology cannot ignore facts about how we reason and arrive at beliefs, the philosophy of action cannot ignore how we arrive at decisions, and value-theory cannot ignore the processes by which evaluations arise. This, of course, is a *philosophical* standpoint, and the argument for this form of engagement might rest on presuppositions others are likely to reject, and for which no further arguments can be offered.

My proposal could be considered as a theoretical shift from a tendency in contemporary meta-ethics: that meta-ethics/value theory is primarily about our *reasons*.[204] I.e. the reasons we *think* in terms of, and use and accept as *justifications*. It is also a turn from the older notion of value as the "fitting object of a pro-attitude", where the relation between value and evaluation is conceived of as somehow *normative*.[205] In general, the reasons, and the reason relation, appealed to in these theories are *normative* as distinct from *explanatory*. This, I'll argue, is *not* the way, or not the only way, that value relates to evaluation. In addition, seeing how the meaning and nature of this "normative" is under discussion, invoking it just moves the problem. The theory defended offers an entirely different sort of explanation: we *should*, in fact, attend to explanatory reasons, rather than to normative reasons, to see what can be made out of it.

[203] Of course, non-naturalist value realist can say that they are true too, but loose out on the other desiderata.
[204] This goes against the current trend to treat reasons as the fundamental normative notion, as *primitives*. See Scanlon (1998), Nagel (1986), Parfit (1984).
[205] See Rabinowicz and Rønnow-Rasmussen (2004).

The point is that there are other, *better*, explanations of our behaviour than the ones we are prone to give.

2.2.5 Descriptivism, goodmakers, and the pattern problem

Naturalism is often construed as a version of *descriptivism,* i.e. as the claim that value predications can be understood as shorthand for descriptions of the valuable objects. *Naturalist* descriptivism is the view that the description required could be carried out in entirely naturalistic, non-normative terms. *Analytical* descriptivism says that this description, in more or less general terms, is accessible by way of a conceptual analysis of 'value'.

One of the most prominent arguments in favour of naturalism starts from the observation that value *depends* on natural properties. The way things are, couched in entirely natural terms, determine the evaluative way things are.[206] Things could not conceivably differ only in evaluative aspects. Value, it is almost universally believed, *supervenes* on natural properties. Admittedly, if non-natural, say *super*-natural, properties exist, value might depend on *those*, but let put this possibility to one side, for now. There are also readings of supervenience that holds that value could not differ *only* in value, while being identical in all *other* respects[207]: This might spell trouble for the naturalist, *if* we want to say that the natural property that value is could, in fact, conceivably vary independently of any other property. The notion of supervenience I have in mind include *identity* as a supervenience relation.

This relates to the fact that we think that people should be able to provide *reasons* for their evaluations in the sense that they must point out why the thing they judge is good is in fact good, whereas some other thing is not.[208] When two similar objects are judged to be of different value, we expect there to be something that distinguishes them from each other that explains the difference.

[206] Jackson (1998).
[207] I'm indebted to Wlodek Rabinowicz for pointing this out.
[208] This does, of course, not mean that this practice is theoretically equivalent to the requirement of supervenience.

In the simplest case, these reasons would appeal to precisely those features of the object that the concept of value should be analysed in terms of, according to the analytical descriptivist. The good object has those properties, the bad, or neutral object doesn't. Unfortunately, most cases appear to be not simple cases. There are a number of complications regarding whether reasons may or may not appeal to features of the person doing the evaluating, rather than of the object evaluated, and whether the features must be *intrinsic* to the object, or whether the value of an object may depend on the relations in which it stands. I will, however, put those matters to aside, for now.

If it is this kind of reasoning that drives the supervenience argument, we find ourselves asking for *relevant* features. When asked to justify an evaluative judgments, it is not sufficient merely to present a description uniquely picking out the thing judged good, such a descriptions can always be rigged seeing how every object has *some* unique characteristic: it must somehow *make sense* of their goodness. To run the argument from reasons parallel to the metaphysical supervenience argument in order to *restrict* what properties could be the supervenience basis is problematic, however. This is, after all, where opinions start to differ. Supervenience in the most general form: i.e. *that* value depends on natural properties, and that *some* reason is required, is supposed to be conceptually true. But what *particular* natural facts make value claims true does not seem to be a conceptual fact. Since supervenience is a necessary relation, the relation between the particular natural facts that determine the value of a thing and that value will be necessary, but that does not make the relation conceptual, or a priori.[209] There must be a *fact* of the matter on which properties value ultimately supervenes, but it need not be a conceptual fact.

The descriptivist argument for naturalism can now proceed in the following fashion: If everyone accepts that value depends on natural properties,

[209] This presents a way for naturalism to preserve proper disagreements in meta-ethics. To some extent, we must agree about what is required in order to disagree about whether those requirements are met or not. Whereas proper, resolvable disagreement would require some common notion of *what* it is that *supervenes,* not merely *that* it does, agreeing about supervenience provide the minimal basis for disagreements: accepting or rejecting the proposed supervenience basis as such.

why not identify value with those natural properties? Even if we don't know precisely *what* those properties are, yet, we could say that value is whatever those properties turn out to be. Frank Jackson expresses this view thus: we *might as well* be descriptivists.

Moral philosophy for most of its history has been in the business of providing general principles for right actions or for evaluating outcomes, so the ambition to find a natural class of properties on which value supervenes is a natural one. But, seeing how most attempts at providing such general principles have been failures; this project has been put into question.[210]

There is an argument from the denial of general moral principles to non-naturalism, recently developed by Graham Oddie[211]: the things we accept as good, and whose goodness we accept to be based on their natural properties, have no natural properties *in common.* Thus there is no natural property to identify as "value". There are two reason why we cannot identify value with the disjunction of (natural) properties acceptable as reasons: First, it would leave us with the problem of how to account for "reasons" in a naturalistic fashion, and second: if we cannot account for "reason" naturalistically, the class of natural properties would be wildly disjunctive.[212]

If we want to argue from supervenience to naturalism, then, we face what has been called the *pattern* problem: there is no pattern to the natural properties on which value supervenes.[213] The point is that while something distinctive is grasped in evaluation, it might not belong on the level of supervenience *basis*: good things have nothing *else* in common. This, if true, would be a reason not to identify value with the (natural) properties found in the supervenience basis.

---

[210] Notably, and programmatically, by particularists like Dancy (1982), Oddie (2005).
[211] Oddie (2005).
[212] Oddie (2005), see Smith (2004) who questions the ban on disjunctive natural properties on conceptual grounds.
[213] Jackson, Pettit and Smith (2000).

A related problem is how we could *learn* basic evaluative concepts if there is no pattern to the things we should recognize as good. Intuitively, we don't think of ourselves as assigning a wildly disjointed property when we say of something that it is good, and we are clearly not pointing to such a complex property when we try to teach a child how to use the word correctly.[214]

If the properties capable of making something good form a wildly disjunctive set, this speaks against the sort of naturalism that would identify goodness with a first-order natural property. The *possibility* of such disjointedness, whether true or not, speaks against accepting that form of naturalism on a *conceptual* basis. The argument for such a form of naturalism, then, should not be a purely conceptual one.[215]

This argument sounds distinctly Moorean: something would be missing from any characterisation of the valuable in purely natural terms. We don't become competent with evaluative language simply by learning to identify a particular set of things, because value consists in something further, something somehow distinctively evaluative in nature. Grasping the concept requires something further. But what *is* lacking? One complaint against naturalism is that judging something to be good is supposed to engage our *motivation* in some manner, and no mere natural categorization would ensure *that*. This idea also drives the suggestion that evaluative terms have an essentially different function from natural terms.

The learning problem could be treated by comparing "good" to a term like "tasty"[216]: You don't learn to use the term "tasty" by learning to identify what people find tasty (even if that is how you are introduced to the notion). You learn to master it by realising the *relational* character of the term: how tastiness

[214] But compare with the way we learn skills like *grammar*: From a restricted sample of sentences we learn how to construct and recognise an infinite variety of highly complex sentences. Appiah (2008) argues that there is a disanalogy because of the *disagreements* in morality. See also de Waal (1996),

[215] Michael Smith (2004) believes that "generalism" is true; good things do have some natural property/ies in common that makes/make them good, but not that it is a *conceptual* truth. Particularists, he argues, are wrong, but not in virtue of getting the *concept* of good wrong.

[216] The analogy is borrowed from Max Kölbel (2003).

depends on *experiencing* things as tasty. The analogy appeals to subjectivists, response-dependency accounts, and possibly to non-cognitivist inclinations (we can easily imagine a theory saying that "tasty" just expresses an emotion) and its aptness is contingent on sympathies with that sort of theory. These theories argue that what is lacking from the naturalist account is an essential relation to motivation, which they find in conceptually relating value to attitudes or to other motivational states. (Response-dependency accounts are naturalist if you allow for natural properties to depend on mental states, see below). As we will see, the naturalist case hangs on being able to find a place for motivation within the theory.

2.2.6 Natural fallacies

In the literature two main and equally serious complaints are lodged against meta-ethical naturalism. The first concerns the lack of a promising candidate property to identify with good, or for the concept of 'goodness' to be analysed in terms of.[217] The second is that, no matter how promising a candidate we can find, "good" cannot be identical to or exhaustively analysed in terms of it. In the literature since Moore, the latter point has been treated as the more important one.[218] The *reason* for this, however, might very well be the persuasiveness of the former point. Some critics have complained that Moore only considered highly implausible naturalist proposals and managed, if at all, to refute only those he did consider.[219] Since one of the positions Moore *did* consider, and at some length, was hedonism, this argument will not help us here.

The open question argument

Moore's "open question argument" famously pointed out that we can always sensibly ask of an object described in any naturalistic fashion, whether it is good

[217] Cf. Sayre-McCord (1988) testability vs. surviving the test thus conceived
[218] With some notable exceptions, like Graham Oddie (2005). Mackie (1977) could be interpreted as offering this sort of complaint.
[219] Notably Jackson (1998).

or not.[220] The point is not that the question has no determinate answer, but the fact that we cannot answer it merely by considering the meaning of the terms shows that 'goodness' does not *mean* the same thing as that description. Note the similarity between this claim and the supervenience argument above: the question whether the value of an object is determined by its natural properties is supposed to be closed, but the question whether an object described in any *particular* way is good remains open. The open question, if it establishes anything, cannot establish the falsity of naturalism in the most general sense, only of particular identifications conceived of as analytically true.

A different reply to the open question argument is to deny its driving intuition: Identity claims can be rather complicated, and interesting enough for it to be doubtful whether they present us with an open question or not.[221] Take a seemingly simple definition like "good = what anyone would desire to desire if free from prejudice and in possession of all the relevant knowledge". It is not immediately obvious that this fails to be an exhaustive analysis of 'good'. Perhaps it does fail, perhaps there are cases where we would like to say that something is good, but does not fit with this description: but that need to be worked out, not taken for granted. It can be an open question whether that question is open.[222]

A related, even more potent argument, treated in greater detail in the next chapter, is the following: Linguistic competence does not necessarily include knowledge of identities.[223] You can be competent with value terms, without being in possession of knowledge about the nature of value or of how the concept should be fully analysed.[224] As long as we have nothing but an intuitive sense of the meaning of 'value', and this sense is a matter of contention, we

[220] The interpretation of this argument is a highly disputed matter. What follows does not take into account every reading of it.
[221] This argument is in Jackson (1998).
[222] However, Wlodek Rabinowicz has pointed out for me that in Moore's sense, the fact that we consider it epistemically possible that something that we would classify as good despite the offered definiens being false *implies* that the question is open.
[223] Lewis (1970, 1972).
[224] See Sturgeon (2005).

can't expect such intuitions to be the last word on evaluative matters. As I will argue, if we do eventually find ourselves in a situation where a notion of value can be offered that mirrors our everyday concept quite closely, and a contrary analysis which offers explanatory cogency, we should opt for the latter. The open question argument, then, does not succeed in ruling out any form of naturalism.

The naturalistic fallacy

Naturalism has been accused of resting on theoretical fallacies.[225] Attempts to identify 'value' or 'rightness' with, or taking evaluative or moral facts to follow from, natural facts somehow gets the problem wrong. In its classical guise the naturalistic fallacy consist in confusing *what* is good, or what *makes* something good, with *goodness itself*. It is an understandable mistake, so the argument goes, seeing how the good (and its good-making characteristics) is what's commonly under our noses when we consider the notion, but to draw the conclusion that to be good just is to be one of those properties is a mistake. The naturalist, then, must either deny that he does make such an identification, or own up it, but deny that it is fallacious.[226] As Frankena argued, "fallacies" seem to be detectable as such only on the basis of a complete and successful theory about the domain. From the viewpoint of a successful argument for an alternative theory, we can identify mistakes in conflicting accounts. But fallacy claims are entirely incapable of providing such arguments on their own: they beg the essential questions. If there is an argument here, it must be found in the reasons behind the challenge, and not in the challenge as such. And to my knowledge an account of those reasons has never been established beyond a reasonable doubt. This is becoming a familiar point: the concept of value is too disputable for any of the contender theories to be ruled out by default.

To treat two distinct properties as if they were one and the same *would* be a fallacy but the distinctive claim in naturalism is precisely that "value" *is* a

[225] Moore (1993), Ewing (1939), Ayer (2001). See Frankena (1958).
[226] This is Frankena's (1958) point..

natural property. Nicholas Sturgeon argues that it would be wrong to identify rightness with something *other* than rightness[227]: i.e. as any natural property specifiable in non-normative terms, but not to identify rightness *as* a natural property. This sort of non-reductionist naturalism holds, I think, little promise, because it fails to connect with the explanatory force that reductive accounts affords.[228]

I will show how a particular natural property, i.e. pleasantness, is prominently featured in an account of a large number of value-relevant facts, and make the argument that this is what needs to be accounted for in a theory of value. When this is said and done, it will be meaningful to ask the decisive question that drives the anti-naturalistic argument, namely whether something have been *left out* of the account. Clearly, as a *reductionist* view, meta-ethical naturalism needs to be put to the test whether it actually delivers what we need. If it doesn't, it might very well be because it is based on a fallacy. This question will, in effect, play the role the open question argument did for Moore in Principia Ethica, but is based on a wider view on how identities can be established.

2.2.7 A hybrid theory of sorts

Consider two theories about value. One analyses 'value' in terms of rational desires, the other cash it out in terms of evaluative experiences: say, that value is what under certain circumstances cause certain "evaluating" emotions. Are these theories obviously incompatible? There is a set of things, objects and actions, such that a rational person (however that is understood) would approve of them. And there is a set of things that cause value experiences, (if there are such things - however they are defined). These sets will overlap to a significant degree, making it hard to decide between them from the standpoint of substantive intuitions. We can undoubtedly find some use for both these concepts, and both are clearly related to many of our beliefs about value. The

[227] Sturgeon (2005).
[228] Sturgeon would surely flinch at this, since he does in fact believe that rightness *is* irreplaceable in some explanations

contrariness of these "theories" is due to the fact that they have their eyes set on the same term, and their specifications include many of the same features. These theories, then, might be found not strictly to be talking about the same thing at all, and consequently not to be inconsistent with each other.[229] The purpose of working this out is to *disambiguate* between concepts of value, if the term can be shown to cover distinct notions. The question then turns to which of these concepts "best" corresponds to, accounts for, or systematizes the features we agree to be relevant to 'value'. Deciding is difficult if we cannot agree about the basic requirements. There are perhaps *pragmatic* grounds for this decision: what concept would be most useful, or would be most likely to receive wide assent? But if "useful" is cashed out, say, in terms of inclusion in the best causal explanation, some questions have probably been begged.

Non-cognitivists might be right about some set of concepts and naturalists about a distinct set, and it can be *unclear* which of these is the "ordinary" concept of "value", or whether there is only one concept at work in the ordinary setting. The task of making this out properly is beyond the scope of this book, but here is an idea about how that story might go[230]: There clearly is room for both the naturalist concept "is desired by me", and the distinct concept that expresses that desire.[231] Because these concepts are so obviously related, and relate to much of the same "stuff" (both involve essential reference to our desires, for instance) they might tend to run together in everyday talk. Both captures something of what we intend with evaluative statements. Thus, for ordinary intents and purposes, the distinction might be unnecessary, even though they are clearly different concepts. There are two ways the story might go from here: Either the concepts cover so much of the same ground that each

[229] See Putnam (1973) on "water" under a different name on twin earth. Of course, we are here considering agents in the *same* world, which complicates matters. I will not go into this here, but I believe that you can treat concepts as covering *contexts* rather than world. See Gibbard (2003).
[230] My supervisor Toni Rønnow-Rasmussen has warned me that this section might rest too heavily on a misunderstanding of noncognitivism.
[231] Note that this latter would be a version of *emotivism*. (Other non-cognitivists say value judgments express prescriptions, or recommendations, or imperatives) Note also that it is only a condition of *sincerity* that the judger actually *has* the emotion; the relation is a matter of semantic *convention*, not of causing the expression. But this is beside the point I'm making here.

of them, with some modifications perhaps, would be sufficient to account for evaluative meaning, making the other superfluous, and it's arbitrary which one. Or, each of these concepts incorporates features that we take to be *essential* to 'value' that the other necessarily lacks, in which case both are needed to exhaust the concept.[232]

To advocate a 'single-minded' theory of value is to ignore what will appear to others as central, and since intentions can be said to rule semantics, to do so would be plain silly. A preferable approach is to restrict one's claim in accordance with what the theory actually covers. I believe that my naturalist theory has some sort of "explanatory primacy" to the non-cognitivist story about evaluations. If there were no value in the naturalist sense, there would be no occasion for expressive terms to carry evaluative meaning. But non-cognitivism might still be essentially *right* about central uses of moral and evaluative language. Since the theory defended here is about *value*, and not primarily about *morality* or even necessarily about *normativity*, we need not to rule out non-cognitivism as an account of (other parts of) normative language.[233] I merely say that value is not exhaustively accountable in such terms: there is an excellent case for value naturalism that such accounts miss out on. Similarly, it would be ridiculous for a naturalist to claim that just because there exists a naturally evaluative property causally responsible for most of our beliefs about value, there is no use for concepts expressing favourable and disfavourable attitudes.[234] No sensible philosopher should claim that the only meaningful statements express descriptive/attributive/referential concepts, nor deny that our evaluative judgments sometimes express those concepts. Our typical evaluations perform a number of mental operations: that seems to be a quite straightforward empirical observation that carries some meta-ethical weight.

---

[232] This might also mean that both should be rejected as possible analyses of "value".
[233] Some theories analyses "value" in normative terms, as, for instance, "being the fitting object of a pro-attitude". If non-cognitivism holds for normativity, then, this would transfer to "value" as well. If this theory is true, "normativity" would seem to be "prior" to "value", since the latter is analysed in terms of the analysis of the former, presumably, does not refer back to the latter.
[234] See Railton (1989).

The disagreement between theoretical alternatives might be based on the conviction that the preferred theory is true at the level that *really* matters. But conflicting theories and the explanatory models they suggest might still be accepted as true as far as they go. Physical explanations are not inconsistent with biological ones, even though they treat overlapping phenomena. If this correctly describes the situation we're in concerning 'value', there might not be any real theoretical disagreement between the main contenders: if there is a conflict, it's about *primacy*, and that might be matter of philosophical taste. As noted, *I* believe that the naturalistic concept has primacy, but, again, this belief is not formed on theoretically neutral grounds.

Given the vague boundaries of what a theory of value should do, conflicting theories could each succeed with *something*: by explaining, enlightening, or somehow accounting for the central intuitions governing this field. Naturalism depends on treating value theory as dealing with a certain set of problems, and on finding them treatable in a certain manner: as we shall see in the upcoming chapter, one such assumption is that "value" can be functionally defined. If these assumptions are declared, it should be clear that they are not *mistakes*, even though they might be *mistaken*. They are justified (in the weak sense that they are *allowed*) by the fact that *no* universally acceptable restrictions on the concept of value bar it from being given a naturalistic treatment. This is not merely to prepare the ground for naturalism (though it is that, to): virtually *no* theoretical approach should be excluded at this stage.

2.2.8 The desiderata

There is not much by way of fundamental theoretical consensus about value. But there is widespread consensus about what value must *approximately* be like, i.e. what the property must be like, if it is a property. If it isn't, the consensus is about roughly what the concept must include, what it must apply to, and what follows from thus applying it. There is some consensus about roughly what a

theory of value has to include in order to be a theory of value at all. It might be that these requirements, as far as they are universally acceptable, are too lax to provide determinate answers, and that a number of theories (nightmarishly: *all* serious proposals) manage to pass the cut. In such a case we need to find other grounds for deciding between the theories in question.[235] Alternatively, the requirements are such that *no* theory succeeds in accounting for value, and value is an illusory problem that disappears after a thorough philosophical treatment.

The criteria people are prone to accept in this matter, or would come up with, if interrogated, are not perfectly overlapping, but they are overlapping nevertheless. That, I've hinted, is why we take ourselves to be *disagreeing* about value. The fact that we differ as to what we find to be *important* criteria, i.e. as non-negotiable features, means that we will not easily satisfy the requirements of those whose basic outlook we disagree with. Perhaps we don't have to. A theory of value, I've said, is a theory that *somehow* explains, accounts for, postulates, fits with, or makes sense of, those things that we do agree upon. There are a number of things we expect to be true about value. These are features that should belong to the property of value if it is to be acceptable as such at all. In this section, I list some of the requirements, or more modestly: desiderata. In the next chapter, we'll see how such a list can be used to shape definitions and prepare the ground for theoretical identifications. The list of desiderata includes, but might not be restricted to, the following[236]:

1. *Motivation*: We expect value to relate to, and be able to engage with, motivation. I take this to be the single most important desideratum. If the normative element of evaluative statements is to reducible to naturalistic terms, surely it must be in motivational terms. A plausible realist theory of value has to show how the property in question engages with our motivational states. This is presumably why desire/preference theories have such a strong appeal, and why hedonism does to. We will eventually have to decide *how* value engages with

[235] These grounds, in turn, will be in need of support, and so on until we're satisfied, or bored.
[236] Similar lists in Peter Railton (1989), Smith (1994) and Jackson/Pettit (1995).

motivation, whether by *driving* it, or by providing *reasons* for it, but this requirement/desiderata does not say which. The theory should pick a side in the old "internalism/externalism" debate: i.e. say whether judgments about value require being motivated to act, or feel, accordingly, or whether the connection is contingent.

2. *Veridicality:* We expect to be right about at least some things about value, i.e. to make at least some correct or near-correct value attributions. We also expect our sensitivity to value, our *judgment,* to improve, if conditions are right, so that we come to make more correct value judgments. Perhaps also that our value-experiences tend to be more on target. We expect to actually, properly, disagree with some people about what is good, and we often expect that most one of the disagreeing parties is right. Some people will be more tolerant to persistent disagreement than others, and some will think it a virtue of theory that it allows value to vary between persons – making "relativity" a separate, although controversial, desideratum, but we don't have to be relativists in order to accommodate a reasonable amount of tolerance for difference.

3. *Fallibility*: Some of our convinced valuations are false: we are capable of realising our mistakes, and correctly identify them as such. Together with requirement 2, this could be construed as the requirement of *independence.* This is to preclude theories that equate value with whatever we happen to value: some of our evaluations are mistakes, and we can come to realise this.

4. *The experience of value*: If there is such a thing as the experience of value, a theory of value should account for it. Whether this experience detects or constitutes value is a delicate question that we'll have to argue about. Is there anything that such an experience detects? What are its conditions, how did it arise?

5. *Supervenience:* We are *committed* to supervenience, to the idea that value depends on natural properties. Jackson (2003) made the further argument that since we accept that value supervenes on descriptive properties, we might as well

be descriptivist.[237] However you feel about *that* argument, you literally *must* admit that natural properties are good-making, and that value necessarily co-varies with them.

6. *Substantive facts:* There are a number of things we believe are good. These beliefs are as firmly grounded as any basic belief about evaluative notions. I.e. whereas we expect fallibility, we don't expect it for central cases: if these are not good according to the offered theory, we might doubt whether we understand the concept, and if *we* don't understand it, it is doubtful whether it is a theory of value at all.

The list might go on, but additions would tend to alienate at least some contestants. We should be prepared to abandon or revise these expectations given sufficiently good reasons, but it is clear that value could not be radically different from what we think it is. It might be different from what most people believe; indeed it might very well *have* to be, seeing how most people are in less than perfect agreement, but it cannot be too different on pain of not being a theory of *value* at all. The list relates to the much shorter list of requirements that Michael Smith called "the moral problem"[238]: how to combine objectivity and motivational force. It's humbling to note that even in this short and supposedly universal formulation, objectivity is a controversial requirement, and motivational force is radically underdetermined by conceptual competence.

Naturalists characteristically say that the things on the list can be accounted for without leaving the natural realm, and it is to this project we will now turn. If something seems to be missing from the theory we offer, we need to ask what that is, whether it can be provided by competing theories, and, ultimately, whether the theory, strictly speaking, is *required* to include it. As not only

---

[237] On the ground that there is no point in adding supervenient properties. On this argument see also Blackburn (1993) who believes non-cognitivism is especially well suited to explain this requirement, seeing how there is no conceptual entailment between descriptive properties and moral predication on most contemporary naturalist views. Even if there is a nomic and metaphysical connection, this doesn't explain the analytical status of supervenience. See Cambell and Woodroow (2003).

[238] Smith (1994).

reductionist, but *revisionist* in character, the account offered will, to echo David Lewis, say that value might not be exactly what we thought.[239]

[239] Lewis (1989).

2.3 Contemporary Naturalism

> *The issue in question must be decided by whatever method we may find satisfactory for determining whether or not a word stands for a characteristic at all, and, if it does, whether or not it stands for a unique characteristic.* Frankena (1939)

2.3.1 A brief history of naturalism

The more or less official history has it that ethical naturalism took a beating in the early twentieth century[240], was largely ignored during the next fifty years or so, due to the dominance of non-cognitivism[241], and then managed to stage a comeback when it was realised that meta-ethics need not be concerned with making analytical claims about the meaning of normative terms. The comeback was largely based on work in philosophical semantics, and referenced heavily work by philosophers like Putnam, Kripke and Lewis.[242] In short, the new wave of naturalists pointed out that true naturalistic identity statements about properties like "good" or "right" are bound to be *interesting* identity-statements. Whereas some of us may think of these identities as obvious on reflection, this is a far cry from claiming them to be a matter of *definition*. Quite the contrary, it is an identity hard won.[243]

---

[240] Influential arguments to this effect were developed by Moore (1993), Ayer (2001), and Ewing (1939). See Sturgeon (2005). They were preceded by Sidgwick (1981). Darwall (2006) points out that Moore's work did not seem revolutionary at the time, and that its widespread influence depended less on these arguments than on the program of philosophical *analysis* on which they rested.

[241] For this version, see, for instance, Foot (1995) and (2001).

[242] Boyd (1988), Railton (1989), Brink (1984), Sturgeon (1985), Copp (1990) More recently: Jackson (1998) and with Pettit (1995), and Smith (1994). The primary texts are Putnam (1973), Kripke (1972) and Lewis (1970, 1972).

[243] This problem is tied to the "paradox of analysis" about which I will, however, say very little

Distinguishing between *a priori* and *a posteriori* necessities, we find that two concepts, F and G, can pick out the same property despite the fact that 'x has F' is not analytically equivalent to 'x has G'.[244] Property identities do not presuppose a definitional equivalence between predicates for those properties, which enable us to defend naturalism against objections based on the lack of meaning-equivalence, like Moore's Open Question Argument.

In fact, it's fair to suspect that the OQA might not rule out meaning-equivalences either, and that, consequentially, naturalists emerge quite unscathed from the encounter. Analytical facts, capturing the true definition of a term, need not be *obvious* facts. One of the two main forms of contemporary naturalism treated in this paper is, in fact, a version of *analytical* naturalism. Even if it doesn't undermine naturalism, the OQA and the scepticism it embodies should still serve as a reminder that identity statements are in need of justification. We need to explain why the identity suggested is not an obvious one, and make the case in favour of it. The quite general possibility to meet the anti-naturalist's objection does not allow us to make any positive claims whatever. The naturalist needs to find and justify other standards for success.

2.3.2 Semantic foundations

One of the main inspirations for the new wave of naturalist theories was the work done in philosophical semantics by Hilary Putnam.[245] Putnam argued that the extension of a term such as 'water' is determined by two factors: the referential intention with which a term is used and the nature of the stuff referred to. For instance, the intention behind uses of the term 'water' plus the facts about the local samples determine that the term 'water' *expresses* the

[244] See Smith (1994). Brink (1989) called the Moorean argument "the Semantic Test for Properties", and claimed that it was not a very good one. I should perhaps add that whereas *this* development was based on work in philosophical semantics, other strands (Hare (1981), Brandt (1998)) were developed from the *irrelevance of linguistic intuitions* (see Ball, 1991). Whether these two lines of thought ultimately come to the same thing is yet another interesting issue I'll not engage with here.

[245] Putnam (1973), See Copp (2000).

property of being H_2O.[246] Putnam suggests this model for the great majority of nouns, and for other parts of speech as well. For a natural kind term like 'water', the term expresses H_2O because being made up of Hydrogen and Oxygen in those proportions is what the relevant samples have in common. To treat 'water' as a natural kind term just *means* looking for such commonalities. It is our intention to refer to this stuff, whatever it turns out to be, which determines the extension of the term – meaning, he concludes, is partly determined by things external to the speaker. Being made of the same stuff is the important similarity between samples of water, but 'importance' in this equation is *interest-relative*.[247] The important similarity between samples is not always sameness of chemical make-up; it might be similarity in origin, structure, function, or what have you.

Taking our cue from this view, then: can 'value' be understood as a term of this kind, and treated in accordance with this model? Assuming that it can: what is the relevant referential intention for 'value', and what constitutes relevant "sameness" for the things we properly assign value to? What is our interest, when we are doing meta-ethics? Might we even say that *naturalism* is guided by a particular interest, such that the best naturalist theory of value need not be the best, say, *practical* theory, i.e. might not be the theory that offers the most straight-forward practical guidance?[248] How do we determine what constitutes relevant sameness? It seems reasonable that the final arbiters of referential intentions are the competent users of evaluative terms. The problem is that competent users (if meta-ethicists are to be trusted as such) disagree on this issue. The matter is complicated further if 'sameness' needs not to be cashed out

[246] This is Copp's way of putting it (2000).
[247] Putnam (1973). For similar arguments about what determines relevant similarity, see McGinn (1976), and Donnellan (1966). Putnam considers the possibility that the meaning of 'water' is constant over possible worlds but that it is world-*relative*, but rejects it on basis of "what we would say" in the hypothetical case considered. The idea that meanings might offer more than one meaning-to-world function has been developed in the field of two-dimensional semantics (See Chalmers, 2006).
[248] Putnam would take this possibility seriously, see his treatment of philosophy of mind in *Reason, Truth and History* (1981). Whereas philosophers like Bernhard Williams would disqualify such a theory precisely on these grounds (Williams, 2006).

in terms of *natural* properties, i.e. properties that play an explanatory role in any science, or, indeed, an explanatory role at all.[249] If 'sameness' is given an unconstrained scope, the outlook for finding a single best realist/naturalist theory of value becomes very bleak indeed. And, remember, this is when we have granted the controversial premise that the relevant intention behind moral language is a *referential* one.

Putnam's theory provides a semantic account for natural kind terms. In the case of the identity of water and H_2O, the most common analogy in this line of reasoning, it seems obvious that some findings would, and did, *count* as identification. Even if this is an example of an a posterori identity, it has been argued, it depends on the a priori status of a more general statement about the success conditions for such identification. [250] The statement that "the substance that accounts for/underlies the "water-phenomena" in this world, is water", according to this argument, is a priori. Can the situation be construed as similar when it comes to matters of value and of morality? In this chapter, we will consider two forms of naturalism, both of which rely on the conviction that meta-ethical claims are roughly analogous to (other) scientific, theoretical claims. This then, clearly, is a good place for non-naturalists to insert an objection.

In his seminal 1988 article "How to be a moral realist" Richard Boyd argued that the analysis of the natures to which natural kind terms refer is not to be conceived as an entirely conceptual affair.[251] Natures are rather discovered via an analysis of a *kind-appropriate* sort: the analysis of the nature of chemical kinds, for instance, is *chemical* analysis, not conceptual analysis. Conceptual analysis certainly plays a part in the methodology for investigating these notions; without it we wouldn't know where to begin, and we would lack the

[249] See Copp (2000). One suggestion is that the relevant sameness is *moral* sameness (Sayre-McCord 1997, Boyd, 2003).
[250] See Horgan and Timmons (1992a and b), for instance. Smith (1994), see below. Chalmers and Jackson (2001) address the more general case.
[251] Boyd (1988), (2003). See below.

resources to rule theories in or out.[252] But, Boyd reminds us; conceptual analysis is an epistemically fallible method, and not the only game in town. We need to engage with conceptual analysis insofar as we must investigate how terms are used and what intentions they are used to express. But we need also to consider what *actually* regulates that use; and this might differ from what we think does. Any such investigation, engaging, as it does, with empirical matters, is unlikely to result in an uncontroversial list of necessary and sufficient conditions. Since value-theory is dealing with an issue of some importance, controversies will run deep and infect this matter no end.

The naturalist needs to say what the meaning of a term 'x' should be like for the property *x* to be a treatable as a natural property, and then show that 'value' meets those conditions.[253] In the last chapter, I argued that if *this* is an open question, then, far from ruling naturalism out, it gives it a fair shot, as it does for *any* plausible attempt. Even versions of naturalism about value that deny the analytical status of the identity need the semantic claim that the term 'value' is *treatable* as a natural property term.

2.3.3 Lewis on theoretical identifications

While Putnam's ideas provided the main inspiration for the first "new wave" of meta-ethical naturalism[254] more recent versions of naturalism owe more to David Lewis work on philosophical method, as promoted in the two texts on theoretical identifications.[255] Lewis describes a route to identifying problematic properties, starting from the "folk-theory" of the relevant domain. Lewis main concern was with *psychological* properties, but the technique is supposed to be

[252] In order to establish that it is a chemical kind we are dealing with, for instance. Notice that "fire" exist as a physical event, but not as defined in folk-lore. The status of identity statements might be discipline relative, or at least *interest* relative.
[253] Notably, this question also depends on whether anything approximately corresponding to the concept thus treated exists. Metaphysical nihilism would *undermine* semantic naturalism, if we are trying to come up with a useful theory, not only with a model for current linguistic practice. See Lewis rebuttal of Mackie (1989).
[254] Sturgeon (1988), Brink (1989), Boyd (1988).
[255] Lewis (1970, 1973). To engage with the *disagreements* between Lewis and Putnam would take us to far. See Lewis.

applicable to all theoretical terms and it seems worthwhile to apply to moral and evaluative terms, too.[256]

The first step in Lewis' account as applied to the moral field is to gather all relevant beliefs surrounding use of moral terms, i.e. the content of "folk-moral theory", and to rewrite them in property-name format. Thus, to judge an act to be right is to judge it to have the property of being right. The second step is to write down a statement conjoining all these rewritten beliefs. These would be roughly the list at the end of chapter 2.2. 'Rightness', thus, is the property that tends to motivate us to act, is favoured by knowledgeable and virtuous observers, befall acts of courage, benevolence, humility, supervene on natural facts, and so on. This goes beyond formulating a *causal function* for moral properties; it says what beliefs should come out roughly true on the theory. The third step is to strip away mention of the problematic property-names, in this case: 'right', and replace it with a free variable. The resulting statement will be a relational predicate 'M' true of the moral properties.

What this predicate get us is a *definition* of the property of being right in terms of the network of relations it stands in to the other properties and facts, moral and non-moral, mentioned in the set of belief: to motivation, action, circumstances of argumentation, acts of an other-regarding kind, and so on. Further, if we replace *all* the moral terms with variables, we get a definition of rightness, and of all moral properties, couched in non-moral terms. On the assumption that the other things mentioned in the platitudes about moral properties are *natural* features, it will be a definition of rightness in entirely naturalistic terms.

The argument consists in two claims: The *conceptual* claim that the right just *is* the x that has these properties and stands in these relations. And the *substantial* claim that natural property F plays this "role". The conclusion to be drawn is that 'right' and property F are the very same property. Provided, perhaps, that

[256] Lewis did so himself (Lewis 1989). I will dispense with the formalization of the theory present in Lewis writings. It is not needed for present purposes.

there exists a unique natural property that relates to other natural properties in the way the conceptual claim tell us that rightness does. If there is no such property, no natural property deserves the name 'rightness'. On the assumption that non-natural properties are ineligible, talk about rightness would then be somehow at fault. This was roughly Mackie's view. There is some leeway, however: We might find a range of "imperfect deservers" of the name; properties that fail to make true all the elements in the relational predicate, but still *enough* of them to be recognisable as *that* property. This, in fact, was roughly Lewis reply to Mackie.[257] Any such departures from the folk-theory of rightness will take some coaxing, however, and will result in a theoretical controversy.

If nothing else, Lewis approach to property-identification provides a way to *diagnose* controversies about terms like 'good' and 'right'. They are disagreements about what a property must do, what must be true about it, in order to be recognised as that property. We might even construe these controversies as a set of theoretical *decisions* about what to require from our theory, about what we are comfortable with as a true referent of a term, and as an account of the property we started out looking for.

Lewis' (1973) model dealt explicitly with theoretical terms definable via a *causal function* that the properties in question were required to perform. Insisting on a similar function for moral terms would beg the question against views that *deny* that these properties have causal powers, never mind views that deny that moral terms refer to *properties* at all.[258] Insisting that the function be *causal*, we should point out[259], is not necessary for the analysis. The folk-theory might include (purportedly) non-causal relations like value being such as to provide *reasons* for actions or attitudes[260]: the nature of the relation between value and favourable

[257] Lewis (1989).
[258] Most sophisticated non-cognitivists thinks that moral terms refer *as well*, it's just that this is not all they do (Gibbard (2003, Blackburn (1993), Hare (1981).
[259] As Wlodek Rabinowicz did to me in private correspondence.
[260] Approximately Ewing's view (1939).

attitudes depends on what platitudes we accept.[261] However, since the naturalist is concerned with getting rid of irreducible normative terms, 'reason' would have to be cashed out in naturalistic terms, or replaced by a free variable in the functional definition.

Even if all normative terms are thus stripped out, the non-naturalist might say, it is still possible that the function picks out a non-natural property. It does, however, mean that this non-natural property would be defined exhaustively in natural terms, so what is there left for "non-natural" to mean? The naturalist claim is that a definition couched entirely in natural terms is *enough* to ensure the unambiguous reference of moral terms. Achieving such unambiguous reference, however, as we shall see, might require the function to have a causal element, even if the folk-theory of the domain doesn't require it. There are further epistemological reasons to promote a causal function for value, which we shall return to later on in this chapter, and in the next.

We need value terms for *something*: there is a "wish-list" of things it would be desirable if the property *did*, and if much of *that* can be captured functionally, Lewis' approach has a lot going for it. If we remove from the list any necessary connection to irreducibly normative notions - those are what we are trying to get rid of, after all - we get a *working definition* for the property set in entirely naturalistic terms.

2.3.4 Network Analyses of Moral Concepts

In *the Moral Problem*, Michael Smith proposes a *network analysis* of moral terms, based on the strategy just outlined.[262] Smith first considers the approach used in the standard view about colour-properties. Colours, on that view, are identical to certain surface reflectance properties. This is clearly not an a priori statement, so how do we arrive at it? What arguably *is* a priori is that colour is

[261] We should, however, also be prepared for the possibility that platitudes, conceived of as universally shared beliefs, *don't say* what kind of relation is required.
[262] Smith (1994).

the property of the perceived objects that causes normal individuals under normal circumstances to have certain experiences. We seem to have nothing more specific in mind when it comes to what qualifies as "colour"; whatever performs that function will do. Identifying colour properties, then, is rather straightforward: you just identify the cause of the relevant experiences. Doing so *counts* as identifying the property, on this model.

Alas, the case is not as straightforward for terms like 'good''.[263] Leaving the question whether we could rely on value-experience to fix the reference of evaluative terms to one side for now: the fact is that it is much harder to come up with uncontroversial conditions for property-identities when it comes to moral/evaluative notions. The general approach might still be applicable, though: There are, after all, conditions that govern how terms are used in evaluative discourse.[264] There is a *folk-theory* of morality, which, in all its imperfections, is where a theory of morality and/or value must begin.

The folk-theory of morality consists in what Smith calls the "platitudes" of moral thinking, i.e. the propositions we hold to be true, and are reluctant to give up on, about value and about the valuable. These beliefs should be incorporated in the identity conditions. Organizing these platitudes into a working definition of the term, we can say that 'rightness' is the property of which the relevant platitudes are, roughly, true. Identifying rightness then becomes the matter of finding what, if anything, they are true about.

[263] And it might not even be that straightforward for 'colour'. See for instance Goldman (1987).
[264] This is perhaps more obviously true for so-called *thick* moral terms, like "courageous" or "humble", i.e. terms that indubitably carries some descriptive content. The network approach, however, is not an attempt to derive all needed content from thick moral notions.

The Platitudes

Smith lists the platitudes approximately as follows:

1) Moral judgments are *practical*. To judge something to be right is, ceteris paribus, to be disposed to do it. Not to be moved to do what you judge to be right requires explanation.[265]
2) Moral judgments are *objective*. By default, a theory of the right should make moral statements objectively true or false. Since the platitudes are revealed not only by our intentions, but in behaviour as well (see below), this is supported by habits of inference. For instance: In moral disagreements, we usually take at most one side to be correct.[266]
3) Rightness *supervenes* on the natural/descriptive: the way things are descriptively determines how they are ethically.[267]
4) The *substance* of morality. We are convinced that some actions are right. We can be undecided on what 'right' means, but be sure that acts of friendship, say, are right.[268] Smith thinks this demonstrates that there is a limit to the content a moral proposition can have if it is to be recognisably *moral* at all.
5) There are *procedures* by which we discover moral truths. We often aim to find agreement in moral matters and more general principles that explain and justify judgments about which we do agree, principles we can then apply to settle disagreements over other cases.

These platitudes, Smith argues, have *prima facie a priori* status. This means simply that they are part of a maximal consistent set of platitudes governing moral terms. The analysis of moral terms consists in, or is derived from, the

[265] Note that for the *meta-ethicist* the reverse is true: it is the fact that morality *does* move you that must be explained.
[266] Non-cognitivists disagree about whether they should give up on objectivity, or somehow accommodate it. See Blackburn (1993).
[267] Jackson (1998), Blackburn (1993).
[268] These might be candidates for "Moorean knowledge", i.e. beliefs that we are more convinced of than we are by the premises of any philosophical argument to the contrary. They cannot support a theory of moral properties on their own, however. To be good cannot be merely to be a member of a group of this kind. (See Moore, 1939).

conjunction of these platitudes.[269] This analysis can be unobvious and informative; it need not be entirely transparent to us what the best summary or systematization of the platitudes is. While mastering a concept, Smith says, consists in mastering the platitudes, you don't have to be able to account for them explicitly. Rather, being competent with the concept of 'rightness', for instance, consists in being disposed to make certain judgments, inferences etc. familiar to moral thought. We might still come to think that these inferences and judgements are *wrong*; our pre-reflective inferential habits are corrigible. But, Smith believes, it would take something like inconsistency for us to give up on such habits.[270] To give up on the platitudes associated with a term is to give up on using the term altogether. Possibly, the relevant beliefs are those that survive in a "mature" version of folk morality, beliefs that we would keep even after facing the relevant evidence and being freed from prejudices and bias.[271]

Smith argues that the moral platitudes can't be captured in a dispositional analysis, such that "good" is simply what causes, and causally regulates, this use of the term. Such an analysis would take the folk-theory as providing a reference-fixing function, rather than as a definition, and it would lead to an objectionable version of meta-ethical relativism: goodness would just be the property that performs the function, and we cannot assume that what does so for our uses of moral terms also regulates them in other communities.[272] This kind of "metaphysical, but not definitional" theory would, he points out, fail to account for why we take ourselves to *disagree* with other people about goodness. One might argue that relativism in general violates the platitude of *objectivity*. But this would be true only if non-relativity does follow from platitudes that should be assigned prima facie a priori status. For relativists, the platitudes

[269] Smith (1994, p 31).
[270] *Or,* I'd say, the existence of an elegant theory that require us to abandon parts of it. Smith does not consider this option. Lewis (1989) does, as do Brandt (1985) and Gibbard (2003).
[271] Perhaps the beliefs should go through something like the "cognitive psychotherapy" Richard Brandt suggested for our desires before they could be trusted as pointing to the good. Brandt (1998).
[272] Smith (1994) p 33. Even if we cannot assume it, we cannot rule it out either. In chapter 2.5, I will make the case that there is such a common cause behind most talk about goodness, which cuts across superficial differences.

about supervenience and objectivity are *conditional* on perspectives, preferences, or what have you, and they would challenge the critic to find the inferential habits that rule out such an interpretation.[273]

Smith argues that if descriptive accounts of the meaning of 'good' and 'right' make our moral judgements inescapably relativistic, and non-descriptive accounts, like expressivism, don't, the latter are preferable.[274] But to say that it is a *tie-breaker* is considerably weaker than saying that it is a *platitude*. The naturalist must find a way to fix the reference of moral terms that prevents moral claims from carrying (too) different contents in different contexts. Yet, Smith argues, relativism seems inevitable if evaluative terms just stand for the properties that causally regulate their use. Naturalists should instead accept some version of *definitional* naturalism in order to secure an *anchoring* function for moral terms.

The discussion illustrates an important general point: if some elements on the list of platitudes are in fact theoretically under-determined by the "folk-theory" of the domain, relativists, or other contenders, can offer an account of the same list of platitudes, just by adding such qualifications. Platitudes, it seems, can be treated by a meta-ethical theory in three ways. 1) Preferably, perhaps; by making them true. 2) By making recognisable qualifications of them true. 3) By offering an explanation of why we might think them true. Relativists will typically try to account for the objectivity intuition in a roundabout way, and show that its falsity follows from (the best systematization of) the other platitudes.

Analysis and the permutation problem

An analysis of moral terms must in some way capture the various platitudes. A definitional naturalistic theory that fails to take this into account will be a

[273] An example is Gibbard's "Plan-laden concepts" (2003).
[274] Hare (1952) rejected descriptivism partly on the grounds that a non-descriptivist (non-cognitivist) account of the meaning or moral judgements better explains moral disagreement.

failure as an analysis. According to the network definition, 'right' is analysed as the property that stands in a certain distinctive network of relations to the natural features mentioned in the platitudes, and to the other moral properties. But if we strip out all mention of moral properties in these platitudes, are we left with enough to uniquely target the moral properties? The risk is that too many properties would fit with the description. Smith calls this the *permutation problem.*

A similar problem arises if we apply the network model to *colours,* for two reasons: First, we acquire colour concepts by being presented with paradigms of the colours. Second, as a consequence, the platitudes surrounding our use of colour terms form an extremely tight-knit and interconnected group. Our colour concepts are not sufficiently defined in terms of relations between colours and things that are *not* colours in order for colour to be reduced in this manner. Consequently, eliminating colour terms is not a good strategy to identify colours. This is not a problem since we have a first hand source of knowledge about colours: what distinguishes red from blue is primarily the nature of the experiences implied, not the relation these colours stand in to other properties.[275]

The platitudes surrounding our normative terms form a similarly tight-knit and interconnected group, so a similar problem arises here. Either we keep a non-reducible moral *residue* to ground the distinctness of the network definition, *or* our definition fails to distinguish correctly the moral/evaluative predicates. Smith believes that the failure of naturalistic definitions throughout the history of moral philosophy is an inductive reason to suspect that such analyses are impossible. Since the plausibility of definitional naturalism is tied to the plausibility of network analyses of moral concepts, this undermines definitional naturalism.

---

[275] Smith does not consider that colour-experiences are, in fact, not colours, and thus that a network analysis of colours would not have to replace names for colour-experiences with variables. The relation between colour and colour-experiences would be problematic only if representationalism about phenomenal content is true, i.e. if the analysis of particular colour-experiences in turn involves essential reference to colours.

2.3.5 Functionalism in ethical theory

Jackson and Pettit offer what they call a *functionalist* theory of evaluative content, and are much more optimistic about the naturalistic program.[276] They argue from two assumptions: the *supervenience* of the moral on the natural and the *networked character* of moral concepts. The supervenience claim, as noted, says that there can be no difference in the distribution of moral properties without a difference in the distribution in natural properties.[277] The distribution of moral properties is *determined by* the distribution of natural properties. Jackson and Pettit writes:

> Characterize a world or an option evaluatively and you assign it to a sort that is adequately identifiable in descriptive terms. Evaluation, to put an ironic twist on the lesson, is description by other means (1995, p 22)

We usually want to be a bit more precise, however: we want to point out *which* properties are the good ones, and know *how* goodness is realised, but the basic claim gives us hope that this could be accomplished.

As for the networked character of moral terms, Jackson & Pettit believe moral terms to be involved in a network of *content-relevant connections* to other terms, including other moral and evaluative terms. 'Fairness', for instance, is tied to 'rightness', so that being fair tends to make an action right. The relations to other nodes in this network are essential to moral terms; there are no plausible *atomistic*, conceptually independent, definitions of these terms.

The interconnections are captured by what Jackson and Pettit call the moral *commonplaces*, equivalent to Smiths *platitudes* and related to my list in section 2.2.[278] They agree with Smith that these are prima facie candidates for *a priori* truths: Anyone who knows how to use the term 'fair' is in a position to see that they hold. Together, these truths give us the conditions under which fairness

[276] Jackson and Pettit (1995).
[277] Non-cognitivists like Hare (1952, 1981) Gibbard (2003) and Blackburn (1993) usually agree. It is a reason to be a naturalist *if realist*. It is not a reason to be a realist tout court.
[278] With the important differences that: 1) I didn't include any normative terms and 2) I talked about 'goodness' or 'value', not 'rightness'.

would be instantiated, while not presupposing that it ever is.[279] Following Lewis, Jackson and Pettit argue that the content of moral terms can be specified by the role they play in the folk-theory of that domain. This theory, we saw, can be scaled down to have a purely descriptive content, by replacing the troublesome terms with variables. As noted, however, the model as such does not exclude the possibility that what this descriptive theory *picks out* is non-natural[280]: the model does not imply naturalism. But the *point* of offering it as an analysis is, precisely, to exclude any mention of non-natural properties: the importance, "essence", if you will, lies in the functional characteristics, not in any particular occupier of the variable space. The suggested approach to content is holistic: the evaluative concepts are not definable on their own. Combined with the argument from supervenience - moral properties necessarily co-vary with natural properties, and these natural properties can fill the role just as well as any "extra" property would - we do arrive at naturalism. But this is, of course, an extra premise.

To apply a moral concept to something is to say that it has the property that plays the role marked out for it in moral thinking. It belongs to certain paradigms that we find saliently similar, it inclines us to judge that an action is correct etc. The term ultimately picks out a natural property not directly, but in virtue of the place it occupies in folk moral theory. Rightness supervenes on/ is identical to the (natural) property that plays the required role. The meaning of moral terms is fixed by the roles described by the commonplaces that cannot be rejected, and would survive in a mature version of folk morality.[281] They are

> the *a priori* compulsory propositions that anyone who knows how to use the terms is in a position to recognize as true. Other commonplaces – other putatively *a priori* propositions – will have to be dismissed as false or downgraded to the status of empirical, contingent truths. (Jackson and Pettit, 1995, p 26)

[279] The substantial platitudes: that some actions *are* fair, would entail that the moral properties were instantiated only if they coincide with the truth of (enough of) the other platitudes.
[280] See van Roojen (1996).
[281] This qualification is more pronounced in Jackson (1998).

This account of moral content fits with what goes on in moral *thinking*, and is similar in spirit, Jackson and Pettit points out, to Rawls view on how our moral beliefs, and beliefs about morality, strive towards a *reflective equilibrium.*[282] Systematic moral thinking involves finding a sustainable compromise between general principles that we find we really cannot give up on, and our considered judgments about how things should be morally characterized.[283] Moral thinking starts from these commonplaces[284], and tries to keep them as fixed as possible when settling on more controversial cases. This might require, as the second sentence in the quote implies, that we give up on some beliefs that we are reluctant to give up on.

Commonplaces about motivation

One feature of particular interest in Jackson and Pettit's account is the required relation of the moral to motivation. The connection works in two ways on this view: First, the property of rightness is defined as engaging with motivation under certain circumstances[285]: to believe that an option is fair is to prefer that option, *ceteris paribus*, to other options.[286]

Naturalism has often been accused of not being able to account for the motivational link, since no natural property engages with motivation in a sufficiently tight manner. Since the precise natures of the platitudes about motivation and motivational strength are contested matters, naturalists can argue that the contingent link that *as a matter of fact* holds between value and motivation, is close enough, considering the theory's proficiency in accounting

[282] Rawls (1971).
[283] Jackson and Pettit (1995, p 26).
[284] Note that this is not an account of how we *acquire* moral concepts. *That* story, of considerable interest in itself, will be appealed to in the next two chapters.
[285] Is it platitudinous that it does so *necessarily*? The existence of externalists seems to suggest otherwise. In general, "folk-theory" does not always suffice to distinguish between theoretically sophisticated options.
[286] Jackson and Pettit (1995, p 23). There are also commonplaces about motivational *strength*: fairness is a stronger motive than politeness, for instance, but a weaker motive than saving an innocent life.

for the rest of the platitudes.[287] The network approach proposes that value should be *defined* as engaging with motivation, which is perfectly compatible with it being a natural property.

Now, if the relation required is irreducibly *normative*, i.e. if the valuable is defined as what *ought* to motivate, this way to pick 'rightness' out is not wholly naturalistic, even though "rightness" could still be a natural property. There would be something normative left unaccounted for in such a 'naturalist' theory. Again: One of the benefits of, and arguments for, the network-analysis is that we get rid of all mention of theoretically troubling terms like 'ought', 'right', 'good', and 'normative': they are replaced by variables. These properties are taken to supervene on natural properties. Now: what matters according to the network approach is that something fills the place cut out for these properties by the folk-theory.[288] If we strip out all mention of normative notions, and allow supervenience to include *identity*, there is no place for a non-natural property that would not also be occupied by a natural property: we have no *need* for non-natural properties in this analysis. Smith's point (above) still stands, however: the definition might not leave us with enough to uniquely identity the right properties, which would be an argument in favour of allowing irreducible normative notions.[289]

Does the analysis ensure that *judgments* about rightness are appropriately connected to motivation? We could make the claim that rightness is the property such that judgments about it tends to motivate, but that would only work when the judgments are *true*, and most people want to say that even *false* moral judgments motivate. This is where the second relation to motivation comes in: Jackson and Pettit suggest that the "canonical" form of making moral judgments is a *non-intellectual* even, partly dispositional in nature: to believe that an act is right, as we said, is, ceteris paribus, to be disposed to do it. Thus

[287] See Railton (1989), and below.
[288] Or by any theory under consideration. It need not be the "folk-theory", but could be the preferred theory of some group of experts, or indeed of any group. We can always stipulate an operational definition, and restrict our ambition to accounting for what that definition entails. In fact, this is not far from what I believe naturalism is, and should be, doing.
[289] Smith (1994), van Roojen (1996).

the link to actual motivation is secured. While we *can* have mere intellectual access to all the relevant moral content, the canonical form, consisting in desires and emotions, provides the essential link between concept and motivation.

What's right?

According to Jackson (1998), the natural properties that occupy the rightness-role form a radically disjointed set: they have nothing *else* in common to which goodness and rightness could be reduced.[290] This is a *substantive* claim: the analysis doesn't rule such a commonality out. It does mean, however, that any such commonality would be coincidental, and irrelevant to the core question of the source of normativity. According to the functionalist view, the "essence" of moral properties is found in the network of relations, not in any particular property that happens to stand in those relations. This is the distinctive claim of this form of analytical naturalism, which puts it apart from the version to be considered next.

Role and realiser properties

If we pursue this approach and define "goodness" as the property of which all the statements listed are true, we find ourselves with two metaphysical options: Goodness is the property that *plays* the goodness-role OR it is the second order, functional, property *to play* this role. The model and the beliefs it incorporates do not say which of these is correct, nor what would decide the matter, so other theoretical and philosophical considerations must be taken into account if this question is to be settled.[291] Jackson argues that since folk theory does not settle this, and no answer is inferable from what it *does* say, the question should be kept open.[292] The difference between the options surfaces when we consider scenarios where something else plays the goodness role from what does so here.

[290] See Oddie (2005) who takes this fact to speak in favour of non-naturalist realism.
[291] As in *wide* Reflective Equilibrium, see Daniels (1979).
[292] Jackson (1998).

Does goodness go with the role, or does it stay with the referent the role picks out in our original context?[293]

To complicate matters, there are two versions of the "role-player" view, too. Let's take 'colour' as an example: Colour is either the property that *actually* causes colour-experiences *or* the property that causes the colour-experiences in the context under appraisal. In the first case, colour terms are *rigid designators* denoting these surface properties even in worlds where they don't, and wouldn't, cause colour-sensations. In the second case colour is the property that causes colour - sensations in the context under consideration. 'Colour', we could say, is *ambivalent* between these two options. If we decide that colour terms designate rigidly, we need a complementary term for the non-rigid concept that allows variation.

The view that Goodness is a role-playing, first order property, picked out by the functional, second-order role derived from the platitudes also comes in two salient varieties: Either goodness is identical to what *actually* occupies the goodness role, even in worlds were it *doesn't* play that role. Or the identity of goodness may vary over worlds, depending on what occupies the goodness role there.[294] Under either construal, the identity between value and some natural property is *a posteriori*. Role-property naturalism, on the other hand: identifying 'good' with the role to play the goodness role, is largely a priori, due to the (prima facie) a prioricity of the role-constitutive beliefs.

Note that non-rigid reference to what plays the goodness role is, for all intents and purposes, equivalent to the claim that 'good' refers rigidly to the *role*.[295] They cover the same cases. The property *to cause x in context y* is certainly distinct from the property *that causes x in context y*, but the distinction

---

[293] If the connection with motivation were *necessary*, goodness would presumably have to go with the role, and not stay with the property that plays this role in the regional context under consideration.

[294] This is the form of relativism found objectionable by Smith above.

[295] Are functional properties candidates for rigidity? I said in chapter 2.2 that if functions are essential, rigidity is not an option. But that holds for what the function *picks out,* not for the function itself. A reason to go for non-rigidity here would be if the function *itself* is not necessary to be recognised as *that* concept. We could recognise a set of functions recognisable as analyses of 'value' or 'rightness', and which particular function is relevant depends on in which world we live. This might be the view put forward by Gibbard (2003).

does not yield distinguishable outcomes in any possible world, so we cannot rely on intuitions about possible world scenarios to decide between them, and it does look like a non-issue.[296]

Rigidity is a semantic/referential relation: it concerns which property is picked out by a term. Rigidity is supposed to capture *necessary identities* in metaphysics. Under the sensible assumption that all identities are necessary, referring rigidly to a role-property lets one aspect of the reference change over possible worlds while the property remains the same, thus saving the integrity of the concept. We don't have to do this, however: there is some plausibility in saying that twin-eartheans have *no* concept of 'water', after all. But that seems to be a rather *pointless* position: If, incongruously, we were to find ourselves on twin earth, thirsting for something to drink, we wouldn't settle for H_2O if it were in solid form, or poisonous: we would be asking after the *functional* kind. The case is even more pressing in the moral case: we *do* want to say how 'good' possible outcomes are, and thus to consider seriously its reference in possible worlds: this is often the basis for our decisions. If there is no agreement about what "kind of kind" we are looking for in meta-ethics, it is probable that this question have no universally acceptable answer.

Possible worlds and Relativity

The case is comparable to the possibility that the water-role be played by something other than H_2O. Since for chemical concepts chemical constitution, not 'function', is essential, 'water' refers rigidly, across possible worlds, to the same stuff no matter how it behaves. In contrast to the thirsting scenario above, we are now asking for water from the *chemist's* point of view. If evaluative concepts are like this, this means that what plays the rightness role here is right everywhere: the possibility imagined is one where things have gone horribly

[296] Jackson (1998, 2003) does not believe in distinct necessarily co-existent properties, like equilaterality and equangularity.in triangles, but it is not clear how he would react to the distinction under consideration here. See also Field (1973), on the indeterminacy of reference.

wrong.[297] But if value is a *functional* property through and through multiple realisability seems to be a virtue, and the possible-world relativism that follows is not obviously objectionable. If we do find it objectionable, we can include an essential relation to our *actual* attitudes in the functional characteristics, so that the property that plays the rightness role in the twin-earth scenario is the property that would have the right kind of relation to *our* values, not to the values of the people in that scenario.[298] Which way to go depends on how we conceive of disagreement with the people in those possible worlds.

If reference can be relative to worlds, it might also be relative to populations. Which the relevant perspective is depends on what fits with the most coherent systematization of our beliefs and intentions. If this holds for moral concepts, we arrive at a version of moral relativism.[299] The best systematization of our moral platitudes might *require* that we give up the belief that moral statements are impersonal and universal. But we might also go the other way: The requirements of objectivity and universal scope might require that we give up, or revise, any other platitude that imply that truth-values are relative to population. If *internalism,* the view that moral beliefs (or "assent to a moral judgment") necessarily motivate action, forces us into relativism, we could consider giving up on internalism instead.

Smith's argument above targeted the possibility that the content of moral concepts varies with groups, cultures or individuals. The scope of our moral claims might be a conceptual matter, but it is a contested one and might even be *vague*; no particular scope is *essential* to moral concepts, just as no particular scope is essential to the concept "we". To repeat: it is not yet clear just what concepts 'right' and 'good' *are*. Or, indeed, whether there is just one set of such concepts. Use could easily be found for more than one. Besides, the alternatives,

[297] The twin-eartheans would, of course, be 'justified' in making the very same claim. This, interestingly, would be no problem if 'justified', too, is restricted to worlds: but note that the normative terms in network definitions might differ in scope.
[298] For extensive discussion of the scope of values and preferences, see Rabinowicz and Österberg (1996), Horgan and Timmons (1991).
[299] See Sayre-McCord (1991) on how relativists could be realists/naturalists.

under the present assumptions, all establish moral properties as *natural* properties, so *this* question can still be considered as settled.

How possible-world cases turn out depends on whether the platitudes invoke essential reference to *actual* attitudes or not, but also on matters of substance.[300] If platitudes are not only broadly functional, but also substantive, this will have an anchoring effect for moral notions: a theory that allows value to deteriorate from certain exemplars would be compromised.[301] This does not mean that the things held as good according to the substantial platitudes would still be good if they did not coincide with the other things on the list: the network definition requires that those things are true *as well*. To be good isn't merely to be one of the good things.[302] It is also probable that reference to substantive values becomes redundant if we include reference to actual attitudes.

Some further considerations

It is not the *nature* of any first order property playing a functional role *that* it plays that role. It might not have done so *in so far as it is that property*. If playing the role is *essential* for what is good, and no first order property does so necessarily, this speaks in favour of identifying goodness with the second order, functional property. Only for this second order property itself is the functional characteristic essential. This is a consideration in favour of role-property naturalism, but there are considerations against it, too. In particular, functionalism has a tendency to turn in to a "what-ever it takes" account, without the capability to *settle* what, in fact, it *would* take, and thus it is powerless to settle disagreements. We will turn to this in the next section.

[300] And vice versa: whether we want to include reference to actual attitudes or not, depends on how we intuitively judge possible-world cases. See Horgan and Timmons (1991).
[301] But see the note on Moore in section 2.1: the meaning of 'value' allows it to belong to anything whatever.
[302] Perhaps the substantive beliefs should be given a restricted scope, whereas more broadly functional features, like the relation to motivation, have a wide, possibly even universal, scope.

2.3.6 Richard Boyd and non-analytical naturalism

Boyd's "How to be a moral realist" is usually considered the canonical statement of the new wave of meta-ethical naturalism.[303] There are considerable similarities between this form of naturalism and the one just reviewed. Both versions put an emphasis on a *role* for moral properties specified by our intentions, practices and dispositions. Both claim that the strength of a naturalist theory depends on its ability to show how natural properties manage with that role. The disagreement concerns the *status* of the role, and what it takes to be a reasonable confirmation of it. The most important difference is that Boyd does not take the specifications of such roles to amount to *analyses* of moral terms. Indeed, he believes that moral terms are not susceptible to analysis in descriptive terms. Instead, moral properties are identified via the *causal role* they play in the regulation of our moral beliefs.

Appeals to semantic facts about 'good' and about other moral terms play a significant role in the defence of meta-ethical naturalism. Nevertheless, Boyd notes, we don't usually think of naturalism in science as the claim that science (except for linguistics, presumably) should be engaged in *language-centred* investigations. Chemists discovered the nature of water without help from metaphysics or philosophy of language. Boyd thinks that since the critique of naturalism is based on the peculiar idea that ethical theory is about language, the criticism becomes void once we put that idea behind us. To be fair, ethical theory *is* involved with philosophy of language; it's just not *all* it is; *that* would be severely and disastrously limiting. Ethical theory involves a trade-off between considerations, and philosophy of language is certainly not barred from contributing such considerations.

[303] At least by Horgan & Timmons (1991) who use it as their main target. This "new wave" of meta-ethical naturalims is also known as "Cornell realism" (See Lenman 2006).

Background

Boyd tells the story of the development in 20th century philosophy of naturalistic, causal, theories of reference[304] and epistemology[305] that tended to treat as *a posteriori* and contingent matters that philosophers formerly thought of as *a priori*; things like the definitions of theoretical terms and natural kinds, and claims about the reliability of senses. Meta-ethical naturalists, he suggests, should make use of these tendencies. Moral terms, much like terms for natural kinds, should be understood as lacking analytic definitions; hence the problem with finding such analyses. Instead, they should be defined in terms of properties and relations that reveal them as suitable for certain kinds of scientific investigation and for practical reasoning. Proposed definitions of such terms are *in principle revisable* in light of evidence or theoretical developments about the relevant properties and relations. This is a very important feature of Boyd's theory, which sets it apart from the view previously considered.

Practitioners within any theoretical or practical discipline, Boyd argues, can be, and often are, wrong about what they accomplish when using terms.[306] Our uncertainties about, and varying uses of, value discourse are fully compatible with these terms referring to natural properties. Not only can I successfully refer without knowing what I refer to, I can do so whilst not knowing that I am *referring* at all.[307] Evaluative practice might, accordingly, support the truth of naturalism even if it is far from our minds that naturalism holds. Indeed, even if we are convinced that it doesn't. The mere *existence* of non-naturalists, in other words, doesn't undermine naturalism (and vice versa). This makes it possible for individuals, or even linguistic communities, who are clearly applying different definitions of moral terms, to refer to the same property.[308] The fact that we disagree about what's good, and about what the function of moral

[304] Kripke (1972), Putnam (1973).
[305] Quine (1974).
[306] Boyd (2003, p 539).
[307] Compare with the non-cognitivist who claims that whereas I might think that I'm referring, I'm (or the "statement" I make is) actually expressing an attitude. Which might come as quite a surprise. Admittedly, most non-cognitivists maintain that moral terms refer in *some* sense, but that this is not the only thing they do.
[308] Boyd (1988, p 195).

terms is, does not rule out the possibility that we are speaking about the very same thing.[309] Indeed, disagreement seems to require this possibility: it is only by presupposing an independent truth of the matter that the notion of a *mistaken* definition makes sense.

This version of naturalism makes the following claim: If any natural property fits reasonably well with the role indicated by our use of an ethical term, *and* is revealed to causally regulate that use, that property is what the term refers to. In order for a theory of a property to be acceptable, and a referent of a term to be found, the property has to be such that the propositions indicated by the word's meaning are at least roughly true about it. This is a common feature between analytical functionalists and Boyd's account.[310] The insistence on a *causal* element is what takes us beyond the analysis offered above. It is not sufficient, or even that important, to make those beliefs *true*: there is the equally important task of *causing* those beliefs, and thus being part of their *explanation*. This feature is supported by the causal, naturalistic conception of reference and of kind definitions mentioned above.[311]

In contrast to the analytical account just considered, Boyd does not take the beliefs exhibited in normal use of moral terms to amount to a *definition* of these terms. The importance of this difference is that the latter theory allows for available explanations of these beliefs to change the working definition of the notion to change to fit with those explanations. If a property is invoked in the causal explanation of why a term is used the way it is, it can earn the status as referent of the term even though it is not immediately picked out by the "a priori truths" about the concept. It is not clear that the analytical approach allows for this explanatory route to identification. Whether it does depends on how strictly 'analysis' is understood.

---

[309] Analogously: two cultures might start out with different definitions of 'water' or 'fire', and yet refer to the same kind.
[310] It also informs Richard Adams' argument in "Finite beings, Infinite goods" (1999) (See Boyd 2003).
[311] Putnam (1973), Kripke (1972), See Copp (2001) for the argument and history of this movement in recent philosophy.

The Pattern Problem

The set of things or properties answering to the concept 'good' defined from folk-theory could be a rather hopelessly disjointed set[312]. The properties that perform a certain function might have nothing *else* in common: the performers are not guaranteed any distinct, natural, unity. Similarly, the function, even if limited as much as plausibility allows, might be suspiciously unstructured, displaying internally quite unrelated propositions. This would amount to a conspicuous lack of *pattern* to the properties, making it unsuitable to treatment by scientific means. Even if it doesn't *in fact* pick out such a set, the method offered does not *guarantee* any pattern to the property. This is an embarrassment for naturalists, and might motivate a turn to other theories, like non-cognitivism or non-naturalism, for the unity we crave for comprehensibility.[313] This is partly for epistemic reasons: How could we *learn* a concept expressing such a disjointed notion? It would still be true that, in Jackson's phrase, we "might as well be descriptivists"[314]: the moral concepts would still cover naturalistically defined sets. But this "naturalism on the cheap" is ultimately unsatisfying. The problem with the analytical view, if taken too literally, is that there is no way to *move on* from a conspicuous lack of pattern.[315] While we do try to keep the content of the concept to a bare minimum, and internal connections between the constituting beliefs are favoured, there is still something arbitrary about this analytical account. Normally, naturalism turns to natural properties because they provide the regularity sought.

---

[312] See Jackson, Pettit and Smith (2000).

[313] Oddie (2005) argues that if the properties picked out by the goodness role display no natural similarity (I wont go into Oddie's definition of that notion here), it follows that goodness is not a natural property.

[314] Jackson (2003).

[315] Kim (1997) argue that functionalism sets moral properties apart from natural properties/kinds in fundamental natural science, because the latter are individuated by causal homogeneity/similarity, and functional properties might be causally heterogenous (Kim, p 301). The functional nature of moral properties is characterized by normative, not scientific, theory. Science only enters to identify the role fillers. While "broadly natural", the properties identified by functionalism allows for no *genuine discoveries*.

Naturalism and Reference

Boyd argues that naturalists should accept two principles about the reference of moral terms. 1) Our use of it must give *epistemic access* to the referent. The referent must be shown to contribute to the regulation of our beliefs so that most of what we predicate of it tends systematically to be at least approximately true of it. This access, and epistemic capacity, may belong to users generally, or to experts to whom they defer. 2) The *achievement explanation condition*: For a term *t* to refer to a property *p,* the epistemic access which uses of t grant to property p must help to *explain* the "theoretical and/or practical successes achieved in the domains of inquiry or of practice to which t-talk is central."[316] This second condition reflects the motivation for acknowledging *a posteriori* definitions in the first place: to explain how *non-nominal* uses of scientific terms, and the practice of metaphysical classification[317], contribute to the explanatory success of science.

The advantage of this approach over the analytical version surfaces in scenarios of the following kind: Let's say we have a natural kind term t, associated with a specified causal role, S. We then find two phenomena p1 and p2, such that p1 more closely satisfies S, but the practices, linguistic and otherwise, of t-users afford them "significant epistemic access" to p2, but not to p1. The causally inclined naturalist can, in contrast to the descriptivist, say that the referent of t is p2, not p1. This, Boyd argues, is a clear advantage.[318]

It's hard to see how this dilemma could arise, however, considering how the role specification is arrived at on a plausible descriptivist view. If it is derived not only from our conscious beliefs but also from dispositions and practices, how could the property *not* be the property we have epistemic access to? The epistemic access condition can even be made a matter of definition, on the analytic view. In fact, if causal regulation is a desirable feature, it should.

[316] Boyd (2003, p 515).
[317] I.e. the attempt to capture *real* rather than *nominal* essences.
[318] Boyd's argument here (as everywhere) is much more prolonged and technical. The account given, I hope, captures the gist of it, and is enough for our purposes.

The analytical view can leave some things to be determined a posteriori. The most plausible version of analytical descriptivism might in fact be a theory for all intents and purposes equivalent to Boyd's theory. Indeed, Jackson suggests giving a modest role to conceptual analysis, not necessarily opposing the causal-historical theory of reference.[319]

Natural kinds

Natural kinds are associated with the notion of causally sustained regularities. The *naturalness* of a natural kind, Boyd argues, is relative to the discipline(s) within which reference to it serves some function. Pain is a natural kind in psychology, but probably not in physics. Natural kind definitions are *accommodated* with empirical conditions, meaning that the world plays as large a legislative role as intentions do when it comes to determining their meaning. Natural kinds are not features of the world *outside* our practice, but of the ways in which such practice interacts with the rest of the world. Accordingly, theories of the nature of the good have the same hypothetical import as theories of the natures of *chemical* kinds when it comes to correctly identifying real essences. Classifications are, to some extent, interest-dependent: the classification into moral properties is conditional on achieving the aim of moral practice.[320] Boyd takes this to show that even if the status of moral properties is somehow dependent on moral norms and practices, investigation into the metaphysics of morals can in principle be carried out by someone entirely *outside* the moral community - all theoretical identifications thus involve a conditional.[321]

Boyd points out that even our best example of an a posteriori identity, water = H_2O, is not as straightforward as we might think. The substance that most

[319]Jackson (1998, p 56) His view can "allow Quine and Putnam much of what they wanted" while letting conceptual analysis seek a priori results. We need just keep in mind that it is a fallible practice.

[320] Sayre-McCord (1997) suggests the relevant discipline is *meta-ethics*, not some *other* science. But we should not prejudge whether meta-ethics is an "autonomous" theoretical domain, and *not* a subdivision of, say, psychology. Copp (2001) argues that it is the business of meta-ethics, not general semantics, to determine the nature of rightness and goodness. This proposal would have more force if we had a clear view of what it is that meta-ethics is *doing*.

[321] Boyd (2003, p 545)

people have epistemic access to is not pure H_2O, but a very dilute carbonic acid. In disregarding this, and making H_2O the referent of 'water', we recognize the importance of the *achievement explanation condition*: it is the connection between uses of "water" and H_2O molecules, and not the (epistemically closer) connection to a dilute carbonic acid that explains how our use of 'water' contribute to explanatory and practical success.[322]

In general, the nature of a thing, property or kind, should be the most fundamental or "metaphysically deepest" candidate available for fulfilling the role associated with the relevant term.[323] This, Boyd thinks, is why we take atomic number to be essential to the nature of an element, and not the set of causal powers summarized by its position in the periodic table. The former is more *fundamental* than the latter because it helps *explain* those powers in a unified way.

The use of *any* term has some causal associations: there are typical causes, mental and otherwise, of our uses of most terms. It would be a mistake to take this to show that causal grounds are essential to the meaning of all terms. Even if we want to resist the claim that the "genealogical" approach is a fallacy in the moral domain, we want to admit that this approach *would* be fallacious for some terms. Not all terms express natural kinds. So, again, what makes a term into a natural kind term? If there is such a thing as the "closest natural kind" for moral terms, we face a decision. It is up to us, the users, to decide whether this is what the term expresses. Is *this* the argument meta-ethical naturalists offer? Being uncertain what type of concept "value" is, we could turn to the closest natural kind and see if we would be comfortable with this kind as capturing the essence of the concept. A complication arises when we notice that, as people tend to assign different weights to the various requirements mentioned, even

[322] Adams (1999), points out that when we consider the possibility that something other than H_2O, say "XYZ", plays the 'water role' in a possible world, H_2O explains what the *role* shared by H_2O and XYZ can't: i.e. the specific features of water and it's behaviour in our context.
[323] How exactly natural properties should be identified would require a "complete" theory about natures in general, and Boyd thinks we lack such a theory.

treating some as non-negotiable, different kinds might be conceived as the "closest" one.

Specifying natures involves identifying what *best fits* the role outlined for a property by ordinary use and practice. This does not require the literal truth of all predications constituting that role. The nature in question can be determined by features of the word's use other than those accessible to conceptual analysis, notably, facts about its causal regulation. Moreover, it requires only that the nature in question be the candidate that *best* fills the role, it need not fit it *exactly*. We are willing to admit mistakes in moral matters, and we are undecided about what is essential to moral and evaluative notions. This speaks in favour of a theory capable of keeping such questions initially open but offering a way to *move forward* on these issues. Meta-ethics is a theoretical endeavour *currently engaged* in the investigation into the nature of the properties that define it; it does not start out from anything like an uncontroversial definition of either its subject matter or its method.

Some people, philosophers very much included, will predictably refuse to budge on any of the requirements. Mackie, as we saw, preferred giving up on ontological realism about value to changing the presumptions that led him to that conclusion. This position is like that of the person refusing to accept the combustion theory of fire since it doesn't account for lightning. Theoretical developments often involve getting rid of preconceptions. And the best way to motivate such a move is to offer an explanation of why those preconceptions are *mis*conceptions.

Concluding remarks

I'm sympathetic to Boyd's views, in particular to the tendency to engage meta-ethics in a wider explanatory project related to the natural sciences, and to allow definitions to be adjusted in light of empirical evidence. I believe this is the sensible route to take for concepts such as "value", i.e. concepts whose content is a contested matter. But it is not clear what we should say about the status of

the following claim: "*If* there is a well-mannered natural property that uniquely plays (enough) of the role specified by ordinary use of the term 'good', that property is goodness." Should *this* be taken as an analytical claim, even on this version of naturalism? The claim, mind, is not that all beliefs constituting the role have a priori status, or are even strictly speaking *true*: it is not a conceptual truth that *any* of the platitudes holds. But would the naturalist not *have* to say that the more general descriptive analysis "to be what, more or less, fills the goodness role" is a conceptual truth about goodness? As we saw above, on sensible readings the versions of naturalism, even though they start out differently, might collapse into the same view.

I disagree with Boyd about the property to be identified, in the end. Boyd thinks that ethical goodness is defined by what he calls a "homeostatic property cluster": a family of properties of actions, policies, character traits etc., which are aspects of, or contribute to, human flourishing and well-being. These exhibit a "homeostatic causal unity", i.e. under suitable conditions they tend to support and reinforce each other. This, Boyd thinks, is the only unity that the good possesses, and thus should be viewed as its *essence* or *nature.* [324,325] Even though the role does not pick out a single, cogent natural property, there might still be a unity to the substantial goods worthy of serving as their nature.

I believe there is a simpler, "deeper", metaphysical fact that explains *more* than would any homeostatic cluster of properties. In particular, it explains *why* and *how* that cluster has initial plausibility as constituting the nature of goodness. The criteria for specifying natures favour such properties, if they can be found. A disjunction offered by the analytical view, or cluster of properties can at best be a second choice. If a property can be invoked in explanations of why such a list includes what it does, or why the properties cluster as they do, it

[324] Boyd (2003), p 510 See Rubin (2008).
[325] In (1988), Boyd wrote that it was supposed to be *conceptual*, *a priori*, what properties belong to the cluster but he seems to have changed his mind (2003). Boyd now thinks kind concepts are *open textured*: there is some indeterminacy in what extension can legitimately be associated with the relevant property-cluster.

is a better candidate for identification. This is the position I will be arguing for in chapter 2.5.

2.3.7 Peter Railton: reducing goodness to happiness

The version of non-analytical naturalism that Peter Railton develops in his (1989) paper is even closer to the view I'll put forward.[326] It starts from the same problem addressed by Smith: moral/evaluative statements seem to have both a descriptive and a prescriptive side. It seems *essential* to statements about the good that they function as recommendations. Indeed, realising this to be true is practically a competency conditions for the concept "good"; not realising that judging something to be good involves recommending it *in some fashion* reveals a deficient grasp of the notion. Accordingly, there are both descriptivist and prescriptivist theories about moral language. It's not difficult, however, to see how a descriptive statement might come to *function* prescriptively, or vice versa. We learn approximately what things people tend to judge as being good, even when we do not agree with the associated sentiment or line of action. If I know what you usually recommend, but disagree with it; I can use your recommendation as mere information-carrying language. All we need for this kind of account is to accept that language can be meaningfully used to express recommendations, and that statements with a descriptivist form are sometimes used in this way. It does not follow that moral terms *primarily* have descriptivist content, nor the other way around.

Railton points out that judgments about the goodness of particular things are often *synthetic* statements, and that such judgments typically concern natural properties, our knowledge of which derives from experience. How does the property of being good fit in to this picture? One option is to deny that 'good' refers (or that statements about goodness have truth values), and thus to say that it *doesn't* fit in to this naturalist account of judgments. They appear to be "synthetic" because they are, in fact, not descriptive statements. The alternative approach, which Railton sets out to explore, is to seek an

[326] Railton does not find this hedonistic program plausible, although he believes in the naturalist project as such (See Railton (1989, 1998, 2000)).

"epistemically respectable explanation of value discourse", and to treat the cognitive, descriptive character of value discourse as essential, and account for the prescriptive force as somehow arising from the substantive content of such judgments.[327] Railton suggests that hedonistic naturalists can meet this challenge.

Railton makes the distinction between *methodological* and *substantive* naturalism that I mentioned in chapter 2.2. Methodological naturalism is the claim that ethical theory is largely an a posteriori matter. It can be conducted as an empirical inquiry, or at least in tandem with abstract and general parts of a broadly empirical inquiry. Railton's primary allegiance is to methodological naturalism, but he does come out in favour of substantive naturalism as well. Moral properties, he claims, are natural properties. Methodological naturalism is guided by the realisation that metaphysically suspect claims are habitually revised or abandoned in light of successful scientific theories. In the process of developing a science, we often find that claims that used to seem true as a matter of logic or conceptual necessity when viewed "purely philosophically", nonetheless seem to be empirically false in the light of explanatorily powerful empirical theories.[328] In retrospect, the *sense* of a priori connections on which conceptual claims are often based, can be explained away. Conceptual categories might present themselves as fixed at any given time, but come to change in accordance with our best theories. Railton suggest that naturalist should be engaged with constructing *reforming definitions*, which are not only revisions of the "ordinary" notions, but also, in turn, *revisable*. Such definitions earn their status by "facilitating the construction of worthwhile theories."[329] The OQA does not apply, directly, to a naturalist theory put forward on these methodological grounds, since such a theory does not pretend to be strictly analytic or even incontestable. Quite the opposite: it is essential to the theory

[327] Others, like Hare (1952, 1993), think this is the wrong side up: you can always change the prescriptive "direction", and then the judgement would no longer be one of moral approval. No actual natural characteristic *entails* prescriptive force.
[328] Railton (1989, p 156).
[329] Railton (1989, p 157).

that the identities proposed by it are a posteriori, and open for revision. Synthetic identity statements can *become* conceptually closed, however, given time and a developing consensus. In value theory/meta-ethics, as amply illustrated by the fundamental disagreements, this has yet to happen.

A naturalist theory of the good, Railton argues, should aim to be revisionistic, and it must establish itself as being *tolerably* so. Reductions/revisions are always in danger of leaving something essential out, and ruling the wrong things in. To meet such objections, and to earn respectability, the theory must shed light upon the function and evolution of the discourse in question. Revisions, Railton adds, can reach a stage were one should say that the concept has been not revised, but abandoned, but there is no sharp line to separate the stages.

Functionalist assumptions

Naturalist about value can use a naturalized epistemology and a natural kind semantics of the kind proposed by Boyd above, to explain our access to value properties, and thus give value a causal explanatory role. Not any causal property will do. It is not enough, for instance, merely to cause beliefs about goodness: Good has a distinctive role in deliberation and action, and the reducing property must be a plausible candidate for this role. In order to achieve a *vindicating*, non-eliminative, reduction, the naturalist must identify good with a natural property that permits one to account for the correlations and truisms associated with 'good'. It must also be able to serve as the basis for the normative function of this term.

Nagel complains that to assume an explanatory role as a test of the reality of values begs the crucial question.[330] To assume that only properties included in the best *causal* theory of the world are real is to assume that there are no irreducibly normative truths. In reply, we should point out that the naturalist perspective is, in fact, *experimental*, and quite deliberately begs the question at

[330] Nagel (1986).

this point. Railton suggests that we see how far we can go, trying to understand this domain of judgment and alleged knowledge by applying a form of inquiry based on empirical models, and to ask how and where these judgments and knowledge-claims fit within the scheme of empirical inquiry. He points out that meta-ethics is at a stage where no theory enjoys much of a consensual backing. The response should not be to give up, but to see how far a naturalistic approach can be carried.[331] Reductions can be vindicating, as well as eliminative.[332]

Reducing 'goodness' to 'happiness'

Railton sets up a 5-step program for successfully reducing goodness to happiness. The first, naturally enough, is to find an *identificatory reduction*: This basically amounts to the claim that happiness is the property that underlies our discourse about a person's good. Now, why say this? It's clear that happiness is one of the things we say make a persons life good, or find desirable for its own sake, but it's not the only thing. Why not say that goodness reduces to, that 'good' *tracks,* some broader constellation of ends? In reply, the hedonist can offer a model for the evolution of desires, more or less equivalent to the conditioning model for desires[333]: Desires that make us act in a way that make us happy, gets reinforced. Most of these desires will have immediate objects other than happiness, and will involve intrinsic interest in other ends, but this origin in happiness is what they have in common.

Next step is the *explanatory role* of the property in question in relation to our beliefs about value. The main reason for reductions in science is the explanations it affords: a reductive theory should explain at least as much as the

[331] Lewis, in a similar spirit, wrote "How much am I claiming? As much as I can get away with." (1989).
[332] If it where not, science would never get of the ground, and we would have to invent new words all the time.
[333] See part 1, particularly the section about pleasure and reward.

theory it is set to replace. An explanatory role is guaranteed if the hedonist is correct about the conditioning model for desires.[334]

The tight, non-conceptual, connection to motivation makes it possible for hedonism to attempt an account of the *normative role* of goodness. The fact that something will make you happy has recommending force, if any simple natural property does. There is a psychological, possibly metaphysical, connection such that we are drawn to happiness.

Is this *revision tolerable*? The theory must be able to capture, directly or indirectly, most of the central intuitions in this area, and lessen the force of those that are not captured. For a hedonistic theory, Railton argues, this indirect route goes via the psychology of desire. The hedonist says other ends owe their hold upon us to the role they play in the creation of happiness. Quite generally, we shouldn't take our theoretically unexamined intuitions at face value. To do so would be to misunderstand the workings of our motivational system.

The task for the hedonist ranges from indirectly capturing intuitive judgments, to explaining those judgements *away.* To explain away intuitions is not always to reveal them as ill-grounded, but to show that they are really about something else[335]: intuitions that appear to be about the good, might in fact track features of related matters like the "right", or the "beautiful". As to the truism that the intrinsic good motivates, the hedonist points out that the experience of happiness is always, necessarily, motivating. There are difficult cases: we might want to recognise the possibility of rational agents who cannot experience happiness, and not all motives seem to automatically generate reasons for action. The hedonist should make a great deal out of the *uniqueness* of pleasure's motivational power. It is distinct from and prior to the derived motivational force of other things. This is why we opt to bypass the "broader constellation of ends" in a fundamental theory of value.

---

[334] The explanation is informative only given a certain conception of happiness, however. If happiness were merely the satisfaction of desire, it would not explain the evolution of desires - for that role depends upon the shaping of desire by the experience of happiness. See part 1.
[335] Brandt's displacing intuitions (1985), Hare on strongly internalised values (1993).

Lastly, the step that Railton calls *vindication upon critical reflection*: Can the reductive account retain its pre-reductive functions, both descriptive and normative? Is the fit reasonably close? Railton says something very clever:

> A closely-fitting reduction might reveal the nature and origin of an area of discourse to be such that we are led to change our views about whether the phenomena to which that discourse purports to refer are genuine, or about whether we are willing to allow the properties which that discourse effectively tracks to regulate our decisions normatively. (p 173)

This is an important test, one that asks us to reconsider the expectations with which we might have entered the theoretical investigation of our concept. Happiness does not "matter" *by definition*; it is just a deep, contingent, fact about us and about the quality of the experience of happiness that it does. The attractiveness of happiness and awfulness of pain underwrites the hedonist's claim. This is argued to be a *sufficiently* tight connection between the descriptive content attributed by this reduction and the "commending force" that accompanies genuine acceptance of a judgment that something is good (for one).[336] One of the main reasons to be a naturalist is that it affords a straightforward epistemic story of the *access* we have to value: knowledge about value becomes possible only when a relevant epistemic capacity is posited.

The revisionist hedonist can, to this extent, accommodate both the descriptive and prescriptive side of value-discourse. The same strategy might be generalized, and applicable to other versions of naturalism as well. Railton points out that naturalist versions of some desire-theories might succeed, due to desire being arguably as tightly connected to motivation as happiness is, and such a connection must surely be present in a plausible naturalist theory trading on causal explanations. In fact, the processes in virtue of which pleasure is a plausible reductive basis for evaluative properties are intimately connected to processes involving desires and thus could be used as arguments for such a theory as well. We will return to this problem in chapter 2.5.

[336] Railton (1989, p 173).

There are philosophers, Hare for one, who oppose the attempt to reduce the prescriptive role of value-discourse in this way: Value discourse is prescriptive whether or not it is causally efficient. But this objection misses the point: Railton offers a *revision* of value-discourse in naturalistic terms. It is an attempt to explain how value-discourse has come to function prescriptively, to offer a story about its evolution. Does this mean that the prescriptive force of value judgments *cannot,* in fact, be reduced, but only explained? As we said, statements about happiness tend to have prescriptive force, and it is very likely that this prescriptive force depends on states of happiness being causally efficacious in motivational processes. But this "dependency" is not one of reduction. As we said above, the naturalist should not deny that there are terms with prescriptive, expressive, meaning, and that our value judgments sometimes, perhaps even most of the time, carry that meaning. If you take this part to be the essential part of evaluative discourse, you are unlikely to be persuaded by the naturalist argument. But there are reasons to resist it to, not least of which are the intuitions in favour of straightforward realism. Revisionist approaches require a decision, to account for some features of ordinary speech derivatively, and take other parts as essential.

2.3.8 The scientific analogy

The analogy between moral properties and natural properties and natural science is at the core of the naturalists' argument, and consequentially a good place to insert an objection. The argument have been put forward that it is not obvious that the analogy holds, which, of course, goes nowhere at all to show that it doesn't. Philosophers like David Barnett (2000) and Stephen Ball (1991) argue that the scientific analogy is question begging, but do not provide an alternative analogy, never mind one which aptness is beyond doubt. The analogy to property reductions in science is valid insofar as it demonstrates that there are cases of a general kind where we accept identity-claims not based on

meaning-equivalences.[337] Presumably, there are analogies for the opposite position as well, a position according to which identities must be fully disclosed to those linguistically competent. For those cases, something like the OQA would be valid. But since it is quite obviously *not* transparent for (all) users of evaluative terms what the conditions for identity are, the latter type of analogy would fall flat. All examples of the non-scientific sort that I can come to think of seem to be cases of philosophically transparent identity statements: transparent for this very reason. 'Value', we have every reason to accept, is not philosophically transparent.[338]

What counts? Trouble on twin-earth

In a number of articles on naturalistic meta-ethics, Horgan and Timmons points to a significant disanalogy between the case of the water = H_2O identity and the naturalist attempt in meta-ethics: It is a priori that certain findings *counts* as identifying water, and this is not so for value.[339] And, presumably, if it is not a priori that something would count as such identification, no such identification can be made. For the naturalist strategy to be applicable, the statement "the stuff that causally regulates our practice with the concept 'good' is good" should be a priori true, and it isn't, so the naturalist strategy isn't applicable.

The naturalist can reply that the role that determines the reference of natural kind terms might actually *not* be known a priori or by conceptual analysis. Kinds can be picked out by features of a concept other than those accessible to traditional conceptual analysis. The role is determined, not by the concepts held by the user, but by the ways in which such use contributes to the

[337] But see Jacksons (1998) claim that Kripke is idiosyncratic in his "meaning of 'meaning'" (1973). I.e. we should not presuppose that science is *not* engaged in a language centred investigation into the meanings of terms.
[338] The suggestion has been made (Rabinowicz, in correspondence) that the Fitting Pro-attitude analysis provides an identity condition that is of the requisite transparent sort. I disagree, insofar as the analysans "ought to be favoured" does *not* strike me as philosophically transparent. Or, for that matter, necessarily true.
[339] Horgan and Timmons (1991), See Rubin (2008).

successes achieved by it.[340] Exactly what would count as identifying fire, for instance, was not always fixed.[341] We should, similarly, accept that it is not a priori that a certain set of findings would count as identifying value. Arriving at it would involve some amount of stipulation. But it should be clear that if a property explaining the troubling features (or some of them) away could be found, naturalist would be satisfied. The analogy to science is useful, not because it is obvious, but because it highlights what this version of naturalism should be thought of as doing.

The argument can offer a different point: even if it was not known *exactly* what it would take to identify 'fire', it was certainly known that *something* would count, it was even known roughly what it would take to correctly identify what fire was. This is what licence treating it as a natural event, and it is the absence of such consensus that undermines the naturalist claim about 'value'. Or, rather: shows it dependence on a controversial premise.

If a naturalistic theory can account for what we actually *do* agree on, there's considerable pressure to accept it. The fact that we agree about roughly what to account for in a theory of value, we said, is what makes conflicting accounts to be engaged in proper disagreement. If this is true, it seems that meta-ethical disagreement concerns what is the best explanation of the things we agree on. The naturalistic hedonistic theory I favour is precisely that: an *explanation* of the features listed. The matter is complicated by the fact that it is not settled that *explanation* is what is called for. But considering the situation we're currently in, if progress is to be made, we need to start down on at least one of these tracks. I'm open to the possibility that conflicting theories might be

[340] In Lenman (2006), the argument from the paradox of analysis suggests that if we have a concept C1 that is murky or problematic, concept C2 might not be, and C2 might be good for all the work we need C1 for. Dropping C1 for C2 might lead to a gain in clarity with no loss in expressive power.

[341] It was *fairly certain*, however, that some phenomena, like burning wood, would have to count as instances of fire: it's hard to imagine this belief coming out false. Hard, but perhaps not impossible.

successful, and think that we should spread our respect liberally in regard to the partial success of each.

None of the arguments presented settles the question what kind 'value' is. The arguments in favour of naturalism, non-cognitivism and non-naturalism seem to depend on presuppositions the others can provide reasons for rejecting. The naturalist might accept this state of affairs and agree that it is by no means obvious that the analogy with scientific identity statements is a valid one, but go on to develop a theory as if it is.[342] I believe, moreover, that a theory starting from this assumption is of considerably greater interest than a theory of the opposite kind, denying the relevance of a posterori findings. Next chapter will be concerned with substantiating that claim.

2.3.9 The Status of Platitudes, Commonplaces and Truisms

We should be prepared to question the status of the "platitudes" or "commonplaces" employed, and to give up on them, if evidence abounds in favour of doing so. I believe that for a number of reasons, we should look for the *causes*, not only the truth-makers, of those beliefs. This takes some justification. To bring about a plausible naturalistic theory requires that we find a way of *accommodating* theoretical needs, empirical observations, and our pre-theoretical dearly held views. There is nothing obviously wrong with the type of functionalism that identifies value with a long conjunction/disjunction of good things, but the theoretical project we are involved in, understood as *fundamental* value theory, can, I believe, do better. The similarities between the things evaluative terms are used to refer to are not as important, or as noticeable, as the similarities in the process by which they get established as valuable. I take this to speak in favour of changing the *focus* from the things usually judged to be good to the properties inherent to this process.[343]

---

[342] This suggestion was made by Brandt (1959), (see Ball 1991).

[343] This feature could be rephrased into a property of the list of things, but it would be in a very forced manner. I will largely forge sophisticated versions of response-dependence accounts, but acknowledge their relevance and status as contenders.

I believe that ground-level, i.e. non—functional, properties suffice to do the required work. If the properties picked out by the function amounts to a long list of properties with no apparent relation (such as being picked out by some attitude), we had better identify goodness with the role-property. But if this is *not* the case, we can identify it with the ground-level property. I think this is what we should strive for, for explanatory reasons: the fact of the matter will demonstrate the common ancestry for what have then developed into quite different practices, thus explaining disagreements.

Possible world scenarios sometimes seem to be the only way to settle these questions, but it seems that intuitions are often guided by what sort of reference relation you take to hold, or what scope you take your claim to have. And this might boil down to what sort of things you believe 'goodness' picks out in *this* world. If you believe it is the typical "things" listed in objective-list theories[344], they are likely to be merely contingent place-holders. Mental-state theories, like hedonism, are much less likely to agree to this. It might be essential to pleasure that it has value, but could it be essential to friendship, knowledge, to works of art? The fact, if it is one, that we would take ourselves to be in disagreement over evaluative issues with people in other possible worlds is a fact about *us*, about our commitment to opinions and beliefs about such matters. It's hard to see how this could establish anything about the metaphysics of the good, however. Would pleasure be good in a functionally turned around world? It would still *feel* good, but, by stipulation, twin-eartheans would not be confused by this. This is hard to imagine, which only testifies to how deeply embedded our phenomenology is with our psychological mechanisms. Strange cases can always be construed to question our commitment to identities: this should come as no shock to us.

When we say of an act that it is right, are we implying that it should be done, or that it has "should be done-ness" somehow built into it? While theoretically important, this distinction would hardly result in a difference in practice. An

[344] See Parfit (1984).

analysis of linguistic competence conditions *underdetermines the content of a theory of value*. To settle for a theory, then, we need to make good on a complementing set of appeals quite besides, and quite possibly even opposing, the linguistic/intuitive ones: we need a theory that fits the phenomena, delivers plausible results in hypothetical cases and has some explanatory power. Metaphysical naturalism has not that much invested in the claim that the concept explained and the concept involved in the explanation must be identical, or even denoting the same property. We are not committed to the view that the "common" sense of 'good' is the most important sense there is. What we are interested in is that there is a property of the designated sort, playing the role specified. Nothing *extra* would be gained by equating it with a vague, general purpose, term such as 'value'.

Few non-naturalists (widely understood) doubt that the *extension* of moral terms cover naturally specified states and actions, what they deny is that this specification captures the *meaning* of those terms, and the essence of the property in question. The approach suggested here is offered as a reply to this claim: it specifies the theoretical requirement for eligibility and then defines the concept of the property to be such as to fit those requirements. Since, by stipulation, nothing but such fulfilment is required, no sense can be made of the claim that the property pointed out is not identical to the property sought.[345] There is still a Socratic possibility: When we have listed the requirements we could find some property that satisfies them that we won't be comfortable calling 'good'. If this is a persistent reaction, the theory dictates extending/adjusting the list of requirements. If that doesn't help, it would seem that whatever that is lacking cannot be captured in functional terms. A rather good argument to that effect is often launched in the philosophy of mind, where so-called *qualia*, so it is claimed, cannot be exhaustively reduced to

[345] Copp (2001) use the "non-philosophical" sense of *meaning* inherent in referential intention to account for the intuitive results presented in moral twin earth scenarios in the papers written by Horgan&Timmons (1991).

functional properties.[346] But surely, this sort of reaction could be undermined given sufficiently good explanations of why we have it. And, so this argument goes, this would not necessarily amount to a debunking of the concept in question, but only of that particular belief.

It seems impossible to settle beyond doubt what semantics fits moral terms, either by mere introspective investigation, conceptual analysis, or any sort of "philosophical" arguments.[347] We've pointed out that this question could be put on hold, to be assessed in retrospect once we've seen what such a theory can accomplish – in light of the a posteriori fact of a property (or cluster of properties) such that there is a credible naturalistic story how the concept 'goodness' came to track it.

[346] See For instance Chalmers (1996), Bengtsson (2003).
[347] Attempts by Kim (1997), Sosa (1997), Rubin (2008).

2.4 The Relevance of Empirical Science to Value Theory

> If this be all, where is his ethics? The position he is maintaining is merely a psychological one. (Moore, 1993)
>
> Ethics must not –indeed cannot – *be* psychology, but it does not follow that ethics should *ignore* psychology. (Stich and Doris, 2006)

2.4.1 Metaethics and the empirical sciences

Roughly fifty years ago, G.E.M. Anscombe stated that

> …it is not profitable for us at present to do moral philosophy; that should be laid aside at any rate until we have an adequate philosophy of psychology, in which we are conspicuously lacking - Anscombe (1958).

Exactly what would be an "adequate" philosophy of psychology is not entirely clear, but the opinion expressed is of interest. The idea that moral philosophy must somehow *wait* for a philosophy of psychology to be completed strikes me as utterly alien to sound scientific practice. Moral philosophy would and should *benefit* from developments in psychology. Considering the development of psychology, particularly in the cognitive and affective sciences since the time of Anscombe's writing, is it now profitable to do moral philosophy *in the light* of this development?

In this chapter, I argue for a theory of value that engages with empirical science. I do so on behalf of the larger naturalist project that I believe meta-ethics should be involved in. Also, because it's a way to move forward given the theoretical standstill we've arrived at using other means. In the theoretical investigation of anything controversial, we should employ every available approach, and this involves empirical research into any potentially relevant

features.[348] This is a theoretical *decision* for which I believe there are plenty of reasons. But they are not *neutral* reasons. I hope to make the alternative position seem less attractive. A meta-ethical theory capable of engaging with empirical science is of considerable greater interest than one to which empirical matters are irrelevant.

The argument for relevance requires a theoretical framework that can translate the empirical element into something that works within the theory. No ethical or meta-ethical conclusions simply *follow* from statements about contingent natural facts; ethical statements cannot be confirmed or disconfirmed empirically without presuppositions about the relationship between ethical statements and statements in other areas of science.

Two things should be noted: First, the fact that no such implication holds for ethical theory in *general* does not mean that no such implications could hold for some *particular* ethical theory, nor that such a theory would thereby by revealed to be false. Second, even if no such *implications* hold, empirical science might be *relevant* for ethics and meta-ethics. We need not be able to strictly *deduce* anything from empirical facts for them to be relevant.[349]

Meta-ethics is not normative

One reason for skepticism towards empirical approaches in ethical theory is that ethics is about *normativity* and normative questions; questions about what we *ought to do* are either 1) not questions about *facts* at all *or* 2) facts of an irreducibly *normative* kind. Now, even if we grant the distinctness of normative matters, this does not bar empirical matters from being relevant to *meta-ethics.* Meta-ethics *is not normative*. It is *about* normativity, and concerns purely factual questions about the status and nature of normativity and normative assertions.

[348] See Katz (1986), Flanagan (1998).
[349] See Greene (2008) Harman (1977, 1986).

A further reason to engage with empirical research is that some of the *platitudes* dealt with refer to empirical hypotheses. Since platitudes on the suggested approach are to be weighed and considered in light of each other, any information we can acquire about them will help. If we are at a tipping point in the disagreement between meta-ethical views, empirical observations might even help tip the scales. They are part of the extended reflective state on which we trade in theorizing.[350] The empirical aspect need not be *essential* in order to be *relevant* in ethical theory.

We also have the theoretical option to *explain platitudes away*. An empirically adequate theory that accounts for why we might *think* that some platitudes hold, while undermining their veracity, might be a needed correlate to an ambitious meta-ethical theory. We should treat platitudes, and intuitions generally, as *corrigible* and *negotiable.*[351] Copp (1990) writes

> Confirmation theory does not *require that the best explanation of our considered judgments be their truth.* (p242))

What meta-ethics?

Appeals to various scientific findings have been made in favour of moral relativism[352], of Non-cognitivism[353] and of error-theory[354]. These positions draw support from *methodological* naturalism. *Substantive* naturalists have been strangely and largely absent from this discussion, this far.[355] This is surprising, considering the emphasis put on the *a posteriori* nature of the identifications suggested in that theory. The investigation into what answers to, or causally regulates, our beliefs about the good is clearly an empirical one. E*ven if we cannot yet say* where exactly our attention should be directed, or what would

[350] "Wide" reflective equilibrium, see Daniels (1979), Tersman (1993).
[351] Railton (1989), Brandt (1998).
[352] Harman (1977), Brandt (1985).
[353] Gibbard (2003).
[354] Joyce (2006, 2007).
[355] Boyd (1988, 2003), whose view clearly favours this sort of collaboration, seems not to have made the effort himself.

"count" as a genuine discovery of a meta-ethical fact there are a large number of question that can be addressed by empirical means.

Philosophical foundations

The idea that theoretical considerations should be sensitive to empirical scrutiny, even though we cannot decide in advance whether any particular empirical fact should be awarded the status of *evidence*, has been a staple in philosophy at least since the writings of Quine.[356] This movement in philosophy points out that beliefs and preconceptions are natural phenomena subject to causal regulation, and in order to properly assess those beliefs, we need to be aware of the processes involved in their formation. Importantly, the processes causally responsible might differ from what we believe to be the basis for those beliefs. This possibility has often been invoked in a *debunking* capacity. Against the notion of philosophy as a "pure" discipline, we can offer a meta-ethical theory *informed* by the empirical sciences. We should do so since the contents of those findings are relevant *anyway,* whether we acknowledge it or not. In assessing the set of beliefs, practices and experiences that make up our ethical competence, we should investigate every aspect of the route by which we reach ethical conclusions and attitudes. We need to turn from a single-minded preoccupation with *justification* to *explanation*. If our beliefs and judgments are better predicted by processes other than what we might think, and unrelated to the *reasons* we tend to give for those beliefs and judgements, these processes are as legitimate objects of theoretical attention as any. A normative account of how our beliefs *should* be formed that doesn't relate to how beliefs are *actually* formed would be incomplete at best, and potentially positively misleading.

The disagreement over the relevance of empirical research to meta-ethics means that its relevance cannot be ruled out (nor in) by default. It also suggests that while it's always an option to *give up* on ever settling theoretical disagreements, we are right to try out all approaches we can to throw some light on the matters

[356] Quine (1974) See Flanagan (1998), Stich and Doris (2006).

under discussion. The position that meta-ethics could somehow be done in conceptual isolation strikes me as way too provincial.[357] This is not to say that philosophical theory should be replaced by empirical psychology; rather, it is to say that the distinction between disciplines was not that strict to begin with.

2.4.2 Moral psychology as part of meta-ethics

Partly for historical reasons[358], meta-ethics has focused on three philosophical areas: epistemology, semantics and metaphysics. Conspicuously absent from this line-up is moral *psychology*. Moral psychology is a discipline in its own right and for the most part it is either excluded from meta-ethics, or *included* only insofar as it pertains to the content of the other three sub-disciplines. There should be little doubt that empirical psychology might be relevant to those areas. *Epistemology*, for instance, if naturalized, has clear affinities with psychology.[359]*Semantics*, if concerned with intentions and common practice, could trade on empirical matters in psychology and anthropology. Even *metaphysics*, if we are concerned with *causal* matters, has something to do with psychological mechanisms. Naturalist leanings in any of those areas translate into naturalist leanings in meta-ethics.

Objections

What case could there possibly be for *not* engaging with psychological theory in meta-ethics? Is it, as Anscombe suggested, the underachievement of psychology in relevant areas? A more general complaint stems from the purely "analytical" approaches to philosophy. Philosophical questions are *conceptual* questions, by definition (what else?) immune to the contingencies studied by observational science. To be fair, the reluctance to accept the relevance of empirical science to meta-ethics might be based on a sound scepticism towards *mistaking* certain meta-ethical questions with other, descriptive questions. It might also hold the empirical approach to a standard: we need to argue carefully for the evidential

[357] See Greene (2008), Katz (1986), Flanagan (1998).
[358] Basically, Moore (1993)
[359] "Why not just see how this construction really proceeds? Why not settle for psychology?" Quine (1974). See Slote (1992).

relation that is supposed to hold, and this might seems to presuppose that a meta-ethical theory is established more or less already. While respecting this striving towards stringency, I don't think it is a profitable way to treat scientific, or indeed philosophical, questions. Again, we should not wait for meta-ethics to make its mind up about how to treat empirical research before we engage with it to see what meta-ethical work it *could* be doing.

The point is sometimes raised that no fact, no observation, could make you believe that, say, torture for its own sake is good. In distinction to our normal, factual, beliefs, our statements on moral matters are not tied to observational confirmation or disconfirmation.[360] This pertains to the relevance of observation to *ethics*, but is held to prove a meta-ethical point: The *reason* why moral statements are not sensitive to observational confirmation is that they belong to a fundamentally different category of statements. Let's see how we could deal with the complaint. It is only to be expected that some natural facts should be *internalised* so that we require a great deal of persuasion to doubt them. While there are probably some outlandish possibilities that would make you believe that oceans are animals, it does in fact appear to be false, and quite self-evidently false in light of available evidence. That some truths are in this way "self-evident" is a far cry from considering them *conceptual,* or otherwise fixed. Such facts, and their status as such, are part of the *explananda* for a natural scientific theory; they are not an excuse from engaging with it.

The irrelevance of pure semantics

We could turn on the argument from empirical irrelevance: if meta-ethics is conceived of as entirely separate from psychology and the rest of empirical science, then *meta-ethics*, not empirical science, is irrelevant to the truly interesting meta-ethical questions. The questions about value that we *can* study using scientific means are of considerable greater interest than those barred from such study. If 'meaning' is disconnected from causal processes, it *doesn't*

[360] See Harman (1977).

matter what 'good' *means;* what matters is how it is used, how we actually function. If there is a conflict between semantic theories one of which *does* engage with actual, causal, processes, we can decide on that as our preferred "meaning". This is the motivation for *operationalisation* in empirical psychology: it means more or less *not caring* about things we cannot operationalise. For example: It's hard to say exactly what *empathy* is, but assigning a specific function to it makes it possible to study, no matter whether the thing studied strictly is what we intend with our ordinary concept of 'empathy'. If we find that something is missing from such an account, we should try to factor the missing part in. If we can't, or if we can't decide *what* to factor in, we should get out of the way. Scientific language strives to dispense with semantic values not open for empirical assessment.

Even if we grant that meta-ethics is exclusively concerned with conceptual truths it should be recognised that the question *what* concept 'value' is might be open to empirical investigation.[361] Once we grasp the concept we might say that nothing further of interest could be found out by observation, but we are not at that stage yet. The fact that we are convinced and somehow competent users of moral and evaluative speech does not imply that we know all there is to know about those concepts. In the last chapter we traded on the questionability of the analytic-synthetic distinction. Even if we honour this distinction, it's unclear on which side ethical identity statements belong. Insofar as meanings are governed by use, it's an empirical question what type of concept "value" is.[362] It could in fact *take* an engagement with empirical matters to establish that empirical science is *not* relevant to meta-ethics, which is it not as paradoxical as it sounds. Mill wrote that

> …when any important consequences seem to follow….from a proposition involved in the meaning of a name, what they really flow from is the tacit assumption of the real existence of the objects so named. (Mill 1950).

[361] See Joyce (2006), Cullity (2006).
[362] Stich and Doris (2006): a "pure", analytical, approach presupposes a meaning already.

This statement, entrenched as it is in Mills general empiricist project, rings true to the approach here. It's not exactly *hostile* to conceptual analysis, but treats it as continuous with questions of fact. Whereas I'm no enemy of conceptual analysis (or only in the sense of being a poor practitioner of it) I think it's insufficient to deal with matters of philosophical controversy. It's useful to make clear the theoretical options available for disambiguation when we find that what we thought was a single concept in fact is a set of overlapping ones. But conceptual analysis cannot *decide* between options; for that we need other criteria and one such criterion, admittedly a controversial one, is empirical adequacy. The identity I will end up claiming for 'goodness' is far too interesting to be directly captured by an analysis: the fact of the matter is simply not accessible by purely conceptual means.

A caveat, however: while it is true that recent advances in neuroscience etc. have made it possible to study processes we had only behavioural and introspective access to before, we should not to be carried away by novelty.[363] Many of the empirical observations which a sensible meta-ethical naturalism can make use of are not new, nor is the employment of them in philosophy.[364] Indeed, philosophy at least up till the "conceptual turn" was filled with reflections and speculation of this kind.[365]

Normativity

An influential objection is that the relation between value and motivation, for instance, is irreducibly *normative* in character. It is not what *actually* motivates that is good, but what *ought* to motivate: To appeal to empirical findings is to commit a category mistake. But we are attempting to construe a theory that *reduces* the normativity of value to natural facts about it. It is not a mere

[363] See Weisberg (2008) on the over-publication of papers with fmri pictures in them

[364] Kahneman and Tversky's research often consists in the experimental testing of common sense hypotheses (2000). If anything has become apparent in psychology in recent years, it is that "common sense" *needs* to be tested.

[365] See Hume (1978), Mill (1993) Epicuros' based his hedonism on psychological assumptions about human nature.

question-begging appeal: it is an explanatory effort to see what, if anything, normativity could be reduced to.

2.4.3 Debunking explanations

Richard Joyce agrees that scientific observations can be of meta-ethical relevance, once the correct domain of inquiry has been identified, but thinks that such observations tend to *undermine* the substantial claims of morality.[366] The best scientific explanation of our evaluative beliefs, language and practices does not refer to the *truth* of those beliefs. Not all causal explanations of beliefs undermine, however. Indeed some beliefs *require* a causal explanation as part of their epistemological pedigree. Empirical findings *could* validate meta-ethical realism, if the facts are right. Harman, famous for making this argument, made a *conditional* claim[367]*: if* there is no reductive account available relating moral facts to natural facts, then they cannot be tested, moral theories cannot be confirmed, and we have no suitable source of evidence.[368] Moral judgment would then be best explained by functions other then their truth.

Personally, I believe that the explanation of relevant phenomena does *not* undermine realism about all values. I *am*, however, debunking *some* apparent values: a veridical empirically grounded theory of value will have to account for the systematic misattribution of evaluative properties. Many, perhaps even most, of our evaluative judgments will turn out to be, strictly speaking, false. But not all of them will.

So what *is* involved in explaining moral/evaluative beliefs? The trick, so to speak, is to look if there is anything present in the causal explanations and behaviour of evaluative judgments and processes that would serve as the "source" of value, and thus as revealing its nature. Now, remember that one of the problems for a naturalist theory is that the things we judge to be good seems to have nothing else in common. Consequently, there is nothing in those

[366] Joyce (2006, 2007).
[367] Harman (1977,1986).
[368] He does, in fact, say that there is such a ground, but that it is *relational.* Of relativism as a form of realism, see Sayre-McCord (1991).

objects that could uniformly *explain* those judgments: there is nothing like the surface reflectance properties in objects judged to be of a particular colour. A source of value would have to be found elsewhere: even if we can still think up a clever way to *locate* value within the object, the explanation will have to turn to the response side to get a distinctive footing.

It has been argued that the particular psychological features involved in the formation of moral judgments are poorly fitted to validate those judgments. Even, that is, if we help ourselves to the notion that moral judgments and evaluations express beliefs, these are governed mainly by emotional manipulations and by self-interest, and by nothing else reliable enough to amount to a proper domain of belief. The best explanation of occurrences of a moral judgment, as well as of the *institution* of morality and value, is one that appeals to functions where the truth of moral beliefs plays no part, and the *function*, social, evolutionary or what have you, does.[369]

Evidence in support of this abounds: It turns out that we are quite bad at judging our own reasons, and factors other than the ones we believe govern our evaluations turn out to outperform them as predictors of judgments and behaviours, suggesting that we could not be trusted to do meta-ethics by appeal to intuitions.[370]

Greene also found moral judgments to be governed by two largely distinct systems, one direct emotional reaction, and one calculating outcomes: selective activation of these two systems accounts for some differences in moral opinion. How to interpret this data is, of course, far from clear, but this knowledge should make us uneasy with a meta-ethical theory that lacks the resources to somehow take it into account.

Other research points out that our *justification* of moral judgments is often an after-thought: we have an emotional reaction, an "evaluative intuition", and

[369] De waal (2006), Joyce (2006).
[370] See Greene (2008), Harman's appeal to experiments in social psychology (1977).

then we try to find the closest, or most flattering, excuse for that intuition being correct.[371] If this was only a tendency, we could agree that we are quite bad at what we do, but that this just means that there is an independent standard for correct judgment. But this reply is not the only one, and it cannot account for the *systematic* nature of these explanations, especially when there exist no other uncontroversial standard for correctness.[372]

We tend to *project* our attitudes onto the world. When an object seems good we take the goodness to reside in *that* object. It is the salient cause, the *occasion*, for the evaluative reaction, and we are disposed to find the explanation for differences in changes in the environment. What we *don't* usually attend to is the fact that reactions are always partly explained by our own internal state. Since our internal state is not that often the object of attention, we disregard its role in the explanation. This is equally true for concepts like 'colour', with the important difference that coloured objects have something in common which appears in the best explanation of our colour-experiences: there is something these experiences can be said to "track". In contrast, the best explanation of our *value*-experiences lies in the internal state of the agent: what we react emotionally to depends on our previous evaluations. In contrast to the basic features of our visual system, our emotional system is severely *plastic,* undermining the hope to find true commonalities in the *object* for those experiences.

Now, one problem with this turn to psychology is that there is a tradition in realism to treat *mind-independence* as a mark of reality for properties.[373] But this need to be understood correctly: the reasonable version of mind-independence is on that claim that properties are not dependent on *being recognised*. *No* reasonable version of mind-independence should rule out the existence of states

[371] Haidt (2001). This is also present in the theory of emotions as *somatic markers*, see Greene andHaidt (2002) and Damasio (1994).
[372] See Kahneman and Tversky (2000).
[373] See Street (2006) and Cullity (2006), who fail to provide a definition of mind-independence that makes it clear that they do not commit this misstake. For a defence of mind-dependence in my sense, see Mendola (1990).

of mind as real properties due to their being dependent on themselves. There is nothing "unreal" about psychological properties, and, as it happens, the explananda for theories of value *must* appeal to psychological phenomena in order to get the required relation to motivation.[374]

Turn from the moral to the good

Most of the contemporary debate about the relevance of science to meta-ethics concerns *moral* matters. As I've said I'm not particularly concerned with morality; my subject matter is *value*, or *goodness*. The empirical study of value has not received the same amount of attention, or at least: the relevance of empirical studies to foundational matters in value-theory has not been assessed. Which is a shame, because it is much more ripe for empirical validation, I think. Value transcends the strictly moral: No particular morality is implied by a theory of value as developed here. Nevertheless, no plausible moral theory could afford to make do without a firmly established notion of value, if one is available. The value domain seems to me "prior", but I won't argue for it further here.

2.4.4 Naturalism and the empirical sciences

The most obvious argument for the relevance of empirical science to meta-ethics follows from a naturalistic *a posteriori* identity claim. We began with a role given by the "platitudes" about value, and then we go on to investigate what in fact plays that role. The identity arrived at is a fact about nature and is to be determined by engaging with the appropriate scientific enquiry, whatever that turns out to be. There is also the related question how we came to have a concept like 'value' in the first place: if it developed out of something else, or by some particular practical need. The best possible confirmation of a naturalistic theory, I take it, would be a "natural" property suitable featured in the explanation of our having the concept in question that *also* fits, to some non-trivial extent, the role specified by that concept. If the best explanation of the

[374] Joyce's (2006, 2007) discussion of the argument from queerness (practical clout) assumes that naturalists deny this requirement. I think this is wrong.

possession and subsequent attribution of a concept features a natural property that fits that concept, we have all the confirmation a naturalist could hope for.

The analytical version

The version of naturalism defended by Jackson says that value is either what plays the role, *or* is identical to the role, defined by the folk-theory of value. This, while sensitive to the empirical question about what is included in folk-theory, is also dependent on the tedious empirical task of identifying what actually performs the role as specified. Tedious, because the role-filler is expected to be a long disjunction of natural properties, but also because it would be like searching for things over three feet tall: The philosophically interesting work is done when empirical matters enter the scene . Analytical versions of naturalism fail, I believe, to latch on to the main reason to be a naturalist at all: the *explanatory potency* of the account offered. By appealing to the platitudes of evaluative/moral thought *directly*, as it were, by making them part of the concept, these versions lose the ability to *explain* these platitudes and their appearing together.[375]

A posteriori identities

The version we are interested in has a more intricate relationship with empirical research. The "role" in question is left vague, waiting for hints from the relevant empirical quarters: if we find a property candidate that accounts for some parts of the role, but perhaps only indirectly or imperfectly for others, this may *count* as finding things *out* about value: we should be willing to revise our concept in light of such findings. There is no *real* issue between the eliminativist and the naturalist approach here.[376]

We do not yet know what properties explain the behaviour of evaluative concepts nor what role it is that the property in question is supposed

[375] Explanatory accounts had a strong standing before the conceptual turn. The dominating status of the conceptual approach depends on what papers you're reading. (See Appiah, 2008) .

[376] As between "no Santa Claus" and "Daddy is Santa Claus". "Strictly speaking" there is no santa claus, but there is an agent causally responsible for the belief. It's just that some (central?) beliefs about it are false. In the Santa Claus case, it's probably correct to eliminate, but this should be decided on a case for case basis.

to play. This allows for genuine discoveries about value: what role defines "value" depends on what is available, what roles are actually being performed. Then we ask if anything looks sufficiently like what we expect value to be. The role might be vague enough to put some doubt whether there is a *property* that plays it at all. "Property work" can be performed by being distributed over processes that require nothing of the kind.

Conditional naturalism

I advocate approaching meta-ethics as a series of theoretical *decisions* about how to deal with the subject matter, roughly conceived. To favour a meta-ethical approach to value is akin to taking a biological approach towards *species*. You do so because biology is an interesting and successful approach to species that seem to reveal important facts about them. One might object that 'species' is a biological *concept* but the fact is that biology, not semantics, made it so. Theoretical intentions, adjustable in light of empirical findings, made 'species' into the concept it is. A similar claim can be made on behalf of the concept of value. The theoretical considerations on which a meta-ethical theory ultimately depends are not further justified. The theory can, in the end, only be tested against the question whether anything of interest has been accomplished, and whether anything essential has been left out.[377]

Some might argue that revisions of the suggested sort means that we are no longer talking about 'value', the property identified would be too far removed from what we believe to hold about vale to be veridical. But, as David Lewis pointed out, in arguments we can distinguish between speaking *strictly* and speaking *loosely*.

> Strictly speaking, Mackie is right: genuine values would have to meet an impossible condition, so it is an error to think there are any. Loosely speaking,

[377] Putnam argued that functionalism is the right *naturalistic* description of the mind/body relation. There might by others, not reducible to the naturalistic world picture. Goodman pointed out that one of the attractive features of non-realism is that it allows the possibility of alternative right versions of the world. (Putnam 1981, p 79).

> the name may go to a claimant that deserves it imperfectly (…)What to make of the situation is mainly a matter of temperament (Lewis, 1989).

Similar sentiments exist in Brandt[378], and it is very much the guiding light of the theory I favour.

2.4.5 Motivation, emotion and proximate mechanisms

Does knowledge of the *evolution* of morality undermine moral realism? It seems that the only reason to believe that evolutionary accounts have a deflationary impact on morality is that there is a further *function* for moral beliefs and discourse, namely a *social* function. Even then, it is conditional on the social function not depending on the veracity of moral beliefs, on recognising moral facts, like the value of other people's lives and experiences. I will not engage with evolutionary accounts here: the science we should engage with is much more "nuts and bolts". We can have no evolutionary account of morality until we have a theory of what the function of morality is and how human beings perform that function. Similarly; there are sound evolutionary explanations of why we need lungs to breathe, but an explanation of breathing need engineering as well: it needs to explain *how* lungs work, not just why we have something that works that way. In short, an evolutionary account of morality and of value would not, I believe, undermine the hedonistic theory proposed here: the function that in fact performs the evolutionary "need" for evaluative practices provides a candidate property whose value is not debunked by its place in evolutionary explanations.

The role of Emotion

Joyce points out that the question *how* natural selection would bring about moral thinking should ideally be answered in neurological and genetic terms. While he cannot offer such an account, he takes solace in the fact that no one else seems able to give such an account at present either. Yet recent empirical research points in a definite direction: emotions play a central role in moral judgments which suggests that if natural selection had a hand in shaping the

[378] Brandt (1985).

human moral sense, the modification of the brain's emotional architecture was its principal means.[379]

I agree with Joyce that the evidence surrounding empirical psychology gives a very coarse-grained answer: that emotion is crucial to evaluation. This, it should be recognised, is not much to go on in a meta-ethical theory. It is not to say that every moral judgment is the product of an emotional episode, or that moral reasoning is always clouded by emotion, or that moral judgment cannot be justified by rational means, or that moral judgments express emotions.[380] There are a number of ways in which emotions can be worked into a meta-ethical theory, and the research so far does not selectively support any particular of them.[381] We should not confuse the fact that moral judgments can be constituents in moral emotions with the view that moral judgments *are* emotions, or *expressions* of emotions. To express an emotion, Joyce points out, is not primarily a causal, but a *conventional* relation. [382] A type of judgment might often emanate from emotions without it being a case of the emotion "expressing itself" through that judgment. Even though no strict inference may be drawn, Joyce does think that the data provide some support for moral *projectivism*.

I'll argue that while our *opinions* about value, what it is and what has it, differ to a worrying extent, the *mechanisms* by which evaluation gets mediated might carry the naturalistic similarities we look for. It suggests that it is not to the intentions of competent speakers we should turn, but to a different level of cognitive functioning. The idea that value is to be found in the object we believe there to be reason to pursue, or in something revealed in the process of conscious reasoning, might be based on a failure to understand how our motivational system works.[383] This charge is, naturally, vacuous until replaced

[379] Joyce (2007, p 123).
[380] Joyce (2007, p125).
[381] Cullity (2006).
[382] Joyce (2008).
[383] Railton (1989).

by an alternative account of how motivation *does* work, and what can be revealed by a thorough study of the matter. It would, regrettably, take us too far to go into the science of motivation in detail, but in the concluding chapter I will point to some research that reveals the role of hedonic processes in motivation. I will then claim that this role gives some weight to the proposition that pleasure can be made out to be a rather fetching referent for the value term.

Motivation

One platitude that gives us licence engage with empirical science is *motivational force*. It is also responsible for the plausibility of moral theories, substantial and 'formal', appealing to *preferences* and *desires*. The platitude is quite unspecific: linguistic competency/folk moral theory alone does not determine whether internalism holds or not, for instance. Nor is it determined whether the motivational force should be transferable through the capacity to recognise *reasons*, and reason-giving characteristics of the object of evaluation.

'Value' must have *some* interesting and fairly close connection to motivation, but given our uncertainty about what connection this is, we need to investigate what connections there actually *are*, and allow for the answer to this question to be something of a surprise, or at least a significant discovery. A property-candidate that not only appeals to, but actually is involved in the *explanation* of something about motivation is better suited for that reason. We cannot say anything informative about how value relates to motivation without knowing how motivation works. Motivation is not uncharted water in the philosophical literature, and its role in moral theory is one of the most covered issues in moral theory. But the focus has been on motivational issues that pertain to the *reasons* people take there to be, both justificatory and explanatory.

Reasons, justification and explanation

While "reason", in its justificatory sense, is a key motivational concept, it is not one I will be concerned with. I must briefly justify (and explain) why: explanation and justifications are related, but distinct, projects and the

naturalistic tendency is to distrust justifications without explanatory force. The reasons why we act, the reasons why we feel and judge, can differ from the reasons there *are* to act, feel and judge. The explanatory reasons can fail to justify and the justificatory reasons can fail to motivate. The reasons we *give* often differ from what actually move us: this is one of the disparaging findings of social psychology, mentioned above.

I think the role of reason-giving in motivation is derivative from the actual causal workings, and that they are to be accounted for. I even recognise this to be a worthwhile philosophical project. But reasons don't *ground* motivation. The mechanisms at work have a structure quite different from the one implied by common reasoning models. When one considers how our cognitive abilities evolved, one becomes less inclined to trust explanatorily impotent accounts.

How emotions work

We are frequently moved by emotions rather than by normative considerations; and by normative considerations only insofar as they interact, at some level, with emotions.[384] Emotions motivate in a way that is often not transparent to us, when we are busy providing the justification that constitute the public sphere of motivation. Emotions motivate in virtue of having hedonic *valence.* They are experience of something as *good* or *bad*. The role of pleasure in motivation is familiar from classical conditioning. Arguably, the motivational force of pleasure is the most basic motivational relation. Objects in the external world get to motivate by engaging with hedonic properties. Even though we seem to be disposed to value some things from birth, what is in fact hard-wired is that these things are *pleasant*. That is what makes us tend, and attend, to it.

The sought *relevant similarity*, the unified natural property that explains the attitudes that ultimately is what makes us think there is such a property as value at all, lies on this level. At later stages in the process, due to the flexibility of the

[384] See Forgas (2004) "Affective influences on attitudes and judgments", Prinz (2006), who favours a sentimentalist version of a response-dependency account for moral facts.

emotional system and changeability of an environment too complex for a pre-set evaluative system, we run across the differences that make for ethical, and meta-ethical, disagreement. The perceived *relativity* of what people care about follows from the process by which we come to learn to appreciate different things.

Empirical psychology has shown people to be prone to make certain projective "errors" when it comes to emotion and evaluation.[385] These "errors" has their explanations: we need external objects, because we need input in order to make the relevant process *start*. We need something that it pays to pay attention to. We tend to attribute value to external objects and their features, which in turn influences not only our substantial intuitions but also our grasp of the evaluative concepts and its nature. This explanation might undermine the concept of value (if we want to keep the features derived from the attribution), but it might also be put in favour of a response-oriented analysis.

This argument is related to the *intrinsicality condition* on early theories of value[386]; the idea that the intrinsic value of an object depends on *its intrinsic features only*, is responsible for a tremendous amount of trouble here. Motivational force can come apart from any object, even when the object seems to provide reasons for the motivational state. Queerness ensues when we try to cram the motivational element into the object for our attitudes. As it happens, the theory I favour does both: it saves the intrinsicality condition while accounting for the motivational link directly. It does so by assigning intrinsic value to motivational states. We will return to this matter in the next chapter.

Some philosophers have proposed *dispositional* theories of moral rightness or non-moral good that make matters of value depend on the affective dispositions of agents. [387] They differ in detail, but Brandt's view is instructive: Ethical justification is a process whereby initial judgments about particular cases and

[385] Related to the "fundamental attribution error", see Haidt (2001).
[386] Moore (1993), Zimmerman (2001).
[387] Harman (1977), Railton (1990) and Lewis (1989), for instance.

general moral principles are revised by being tested against the attitudes, feelings, or emotions that would emerge under appropriately idealized circumstances. The "test" of moral values consist of the attitudes people would have if they were impartial, fully informed and vividly aware, and free from abnormal states of mind. Actual individual differences pose no threat to objectivity if such a universal standard can be found.[388] But would everyone have the same attitudes in ideal circumstances, so that they would eventually *settle* on certain objects? Brandt turned to anthropology and found that there are disagreements that seem impossible to track down to difference in non-moral beliefs. He ultimately held that the justification he had in mind ends in relativism, since "groups do sometimes make divergent appraisals when they have identical beliefs about the objects."[389]

The non-relativist realists need to explain away such failures of convergence in a way consistent with the objectivity of moral judgment and rational resolvability of moral argument. A turn to the response-side, again, might be the only possible way to find such common ground. It is, as we shall see in the next chapter where I develop the hedonic explanation of value, to such a response we should turn.

[388] Compare with Hume on the matter of taste (1874-5).
[389] Brandt (1954, p 130)

2.5 Naturalist Hedonism

You can't be good until you've had a little happiness
E.M. Forster, the longest journey

2.5.1 Introduction

In this chapter, I develop the hedonistic theory of value the last four chapters have been spent in preparation for. The explanatory approach, and the empirically informed version of metaphysical naturalism, supports a version of hedonism. The argument depends on the role pleasure plays in the formation and regulation of evaluative beliefs and practices, and its role in the explanation of processes relevant to our beliefs about value. An explanation in terms of hedonic processes should, I think, be accepted as vindicating, rather than undermining, evaluative beliefs.[390] What is true about pleasure is *sufficiently close* to what we hold to be true about value for pleasure to be a suitable referent of that term.[391] Still, the theory has the result that many of our evaluative beliefs are false, indeed systematically false, and this it needs to compensate for by explaining why we have those false beliefs.

It seems that most people at some point in their thinking life entertain hedonistic inclinations, and a great many philosophers have defended this view. Why is that? Exploring how pleasure relates to what we believe about value will, minimally, help us to account for this inclination. What I hope to show is that the basis of this inclination can be turned into a plausible theory of value, that the inclination is, in fact, a truth-tracking one.

[390] Copp (1990), Railton (1989) Joyce (2006, 2007), Harman (1986).
[391] The "sufficiently close" argument is made by Lewis (1989), Brandt (1976), Railton (1989) and is mocked by Joyce (2006).

The position I defend is not merely that pleasure is valuable.[392] Pleasure has a special relation to value, and the naturalistic framework lets us claim that this relation is the most intimate possible: Goodness and "pleasantness" – or whatever one should call the property that makes pleasure into what it is - is the very same property.

In the last chapter we addressed the question whether value theory has or should have any empirical commitments, and how value-theory could be made an empirically responsible discipline by engaging with matters of explanation, thereby making it continuous with a more general quest for knowledge that includes the empirical sciences. Hedonism has everything to gain by being involved in such an explanatory project.

Here's a very straightforward kind of explanation, along these lines: Pleasure (and pain) causally influence our first-order evaluations. Thus, if pleasure and value is the very same thing, value figures prominently in the explanation of our evaluations and evaluative beliefs. In particular: pleasure causes the belief that pleasure is good. A similar story tells how we come to believe that other things are good, but since pleasure uniquely plays this role, those beliefs are *not* vindicated by this explanation, whereas the role of pleasure in those explanations shows the "depth" of the truth of hedonism: the fact that pleasure is more than just *a* good: it is the origin of our evaluative beliefs. This is the core of the argument, and we should not be too quick about it. The centrality of pleasure in the relevant explanation does not need to translate into the view that pleasure is the good. If we didn't believe that pleasure was good, its role in the relevant explanations could show one of three things: 1) There is no such thing as value. We have a full explanation of our beliefs about value that does not include any reference to it. Or 2) Value is somehow *picked out* by pleasure, but does not belong to it, or not only to it. 'Value' could be understood as a

[392] Very few non-hedonists deny that pleasure is good. Among them, some claim that *all* pleasures are good, but that other things are good as well,, whereas others thinks that it is only some pleasures that are valuable. Fred Feldman recently suggested that a theory qualifies as hedonistic if it claims that pleasure is the only thing that has positive intrinsic value. Feldman (2004).

response-dependent property, and pleasure as the relevant response.[393] Or 3) Pleasure could have an *epistemological* role, being something like the "perception" of value, a value that is existentially and perhaps conceptually independent of this act. While these interpretations are worthy of consideration, I will argue for identifying pleasure with value, rather than for giving it a derivative role in the theory of value. For one thing, if the sort of objects typically associated with pleasure were to come apart from it, I believe the value most plausibly would go with the pleasure, not with the objects.[394]

Four pillars of hedonism

There are roughly four main pillars of the theory:

First: *Motivation*. The most important argument in favour of hedonism comes from the connections that hold between pleasure and *motivation*. This was Mill's point in "Utilitarianism"[395], and it is, with some qualifications, my argument here.[396] In order for a property to be acceptable as value, it must have a non-trivial connection to motivation. But the nature of this connection is not settled by the common conception of value. The approach I favour holds that this connection is one of the things we need to *find out*. The inquiry into the functional and experiential characteristics of pleasure reveals that it plays a very central role in motivation indeed.

Second: *The common element*. Hedonism claims that there is a natural property that all intrinsically valuable things have in common, and that serves as their "essence". Admittedly, we arrive at this commonality by denying the (intrinsic) value of many things normally judged to be valuable, and this takes some justification. We need to explain how we came to hold those beliefs, and how a

[393] Variations of this view have often been proposed in the field of aesthetics (Hume (1874-5), Marshall (1892 and1893) Goldman (1990), Petts (2000), see Brax (2009).
[394] Notoriously, in cases were our *economic* behaviour comes a part from our happiness, the common inclination is to judge that we should follow the happiness, not the money.
[395] See Mill (1993), Criticism in Moore (1993), but see Millgram (2000) for a defence. Moore admitted that there is a quite intimate relation between pleasure and motivation, but did not think it was the right one.
[396] Similar arguments have been proposed by Mendola (1990, 2006), Katz (1986, 2008), Crisp (2006), Railton (1989).

term that denotes a simple natural property came to have such a great array of uses.

Third: *Pleasures are Value Experiences*. A theory gets its character by the data it presumes to account for, and there quite clearly are such things as value data.[397] I've mentioned value judgements, reasoning, behaviour and the rest of the beliefs and dispositions that make up the platitudes surrounding value. But there are also value *experiences*. Value experiences often prompt, or accompany, or confirm value judgements, but it is far from clear how they should fit within a theory. The sorts of experiences that could serve this purpose are the experience of Desire, Emotion and Pleasure. I hold that the core evaluating element in all of these candidates consist in their affective character. Pleasure just *is* the value experience.[398]

Fourth: Hedonism offers an approach to the *epistemology* of value. In fact, it offers two forms of epistemological access to value. First, if pleasure is value, the experience of value affords direct, "phenomenological", access to value. Second, it provides a causal route from value facts to value beliefs.[399] This is one of the motivations for the empirical project and, I'll argue, it is in this way that we can demonstrate how value fits in the explanatory picture of the world.

Hedonism is a contested theory, and a number of objections have been lodged against it. At the end of this chapter, I will show how the theory developed can be used to answer those objections.

---

[397] See for instance C.I. Lewis,(1947) Oddie (2005), Mendola (1990).

[398] I sometimes slip into using the phrase "the experience of value" instead, which unfortunately seem to imply that there is a distinct value that pleasure is the experience of. As I say, it often seems that way, as we tend to assign the value experience to the object of that experience. But the object of the experience rarely has the explanatory potential required to play the part.

[399] One sought, and not found, by Harman (2000) and Joyce (2006). Railton (1998) points out that a theory of value needs to account for both semantic and epistemic access to value. He points out that moral claims play a role in action, in judgment and in society that would not be possible if they were not "accessible through experience and intimately connected to the *how's* and *why's* of human conduct." (p175-6).

2.5.2 Naturalistic Hedonism

The distinctive feature of hedonism, understood broadly, is that it gives *pleasure* a central role in the theory. According to *psychological* hedonism, that role is in human motivation.[400] According to *ethical* hedonism, the right action is the action that maximize pleasure, or the action that is intended to do so. Most commonly, hedonism is offered as a theory of *well being* – about what makes a life good for someone. It then becomes a theory of the *good* if the good is reduced to matters of wellbeing.[401] I will defend hedonism primarily as a theory of the good, a view we can call *axiological* hedonism.

Each of these positions can be held separately, or in conjunction and they can be used in arguments for each other. I believe that there is a strong case for identifying pleasure with the good, and that this is supported by a version of psychological hedonism. But I'm not committed to the *moral* version. I tend to think that morality has something to do with making things better or worse - we ought to aim at the former and avoid the latter - and I think better and worse should be cashed out in hedonic terms, but I don't think there is any single way of doing this. In fact, the theory I offer suggests that there are several.

Hedonism about the good says that pleasure is the only thing that is good in itself. A *naturalistic* version of this claim is even stronger: Pleasure is what good *is.*[402] This is to be contrasted with the view that pleasure is good *because of* some distinct property it has. For instance, the view that the good is what we would desire if we were fully rational and that, as a matter of fact, pleasure is what we would desire if we were fully rational. During the last hundred to hundred and fifty years or so, hedonism has primarily been proposed as a *substantive* view,

[400] Whereas the label "psychological hedonism" usually denotes a specific version of this claim, it should be taken to cover a wide range of possibilities.
[401] See Sumner (1996), Griffin (1986), Nozick (1974), Mendola (2006), Parfit (1984), Crisp (2006).
[402] In terms recently employed by Roger Crisp (2006), it is both *enumerative* and *explanatory* hedonism. The former is just the claim that the list of good things includes pleasures only, whereas the latter is more accurately expressed as the claim that pleasure *makes* things good. Similar proposals appear in Railton (1989), and Katz (1986).

saying that the list of intrinsic goods consists of pleasures only.[403] A naturalist version claims that pleasure and value are the very same thing. There are reasons to believe that the hedonists of old intended something along these lines, though we should keep in mind that the clear distinction between these forms of hedonism is a rather modern invention.[404]

Functionalist hedonism

If we adhere to the broadly functionalist approach dealt with in chapter 2.3, Naturalist Hedonism can take at least two forms: 1) Value is a functional, second-order property, and pleasure is the property that in fact plays that role. Or 2) Value is the property *that* plays the role just mentioned, and, again, pleasure is the very same property. The first option is a version of naturalism that is only contingently hedonistic: value is a natural property that just happens to belong to pleasure. The latter option is the more ambitious claim for the hedonist, and also a better approach for it. It says that pleasure would be good even if it *didn't* perform the role it actually does perform. It would then be a standard natural kind identity statement: necessarily, but not conceptually, true. In chapter 2.3 I argued that if the platitudes don't pick out any single natural property, or if the set of properties picked out has no other common feature, we would be better off identifying value with the value-role. That's what those otherwise diverse properties had in common. But if there is some single property that stands out as what the platitudes ultimately track, we should go for that property instead. For one thing, it would furnish us with a *foundation* of the platitudes, against which they might then be compared and judged as trustworthy or misleading.

But we also encountered considerations in favour of the role-property version of naturalism. If we take the functional characteristic to be *essential* to value, the possibility of a separation between a candidate property and this function would disqualify it out of hand. It is often required that value

[403] Arguably, this was Sidgwick's doing

[404] Bentham (1789, 1960), Mill (1993), Epicuros. Moore writes that hedonism "appear in the main to be a form of Naturalistic Ethics", by which he means that it's held to be the sole good because it "seemed somehow involved in the *definition* of 'good'" (1993 , p111).

somehow be imbued with motivational power, for instance. If this can be included in the role, but does not belong essentially to any plausible role-filler, we should opt for the role-property version. But if some first-order property actually does come with a suitable motivational impact in all possible worlds, this consideration no longer favours role-property naturalism. The argument for making the direct identity claim depends on two central claims: First, that we should treat the *phenomenology* of value as essential to it. The second is based on an explanatory tactic that has more in common with Boyd's causal version of naturalism. I will begin with the latter.

Causal Hedonism

The form of naturalism I favour makes a claim about how the concept of value can be treated, but it is important to note that the hedonist identification is not an analytical claim about the meaning of 'value'. It is an explanatory account, making a metaphysical claim based on the fact that hedonic processes causally regulate our beliefs about value. While I believe that pleasure actually performs central parts of the value role as defined by the platitudes – in fact, *enough* of it – the case for hedonism, as for naturalism in general, depends on the fact that pleasure is a key part of the *best explanation* of this role. It is part of the best explanation of why the platitudes are what they are, why they cluster the way they do, and also of why their relative strengths and perhaps even their content, seem to differ between persons and cultures. Offering such an explanation is something hedonism *affords*. This is something we couldn't do if value were *defined* as that role. Value can be invoked in explanations of motivation in a way that it couldn't if value was motivating by definition. In Boyd's terms, pleasure is the metaphysically "deeper" fact about value. Pleasure is not good because it occupies a certain role. It is *recognisable* as the only good for that reason, but its goodness does not consist in/ is not constituted by playing that role or being picked out in that fashion.

The main differences between Boyd's version of naturalism and the hedonic version I favour are that 1) I take value to be a single, simple natural property,

and not a "homeostatic cluster" of properties. Interestingly, Boyd's own criteria favour simplicity in the property candidate, yet he does not consider hedonism as such a candidate. 2) The property that causally regulates our beliefs about value does not belong to many of the things we typically find valuable. Most of our ordinary value-attributions are false. These two points are, of course, related: it is precisely because I settle for a single, simple natural property that such a large number of our ordinary evaluative statements turn out false. Where the first point seems to be a considerable advantage, the second is something of a theoretical burden.

The basic claim

What does it mean to say that pleasure is good? Is it, as Moore famously argued it couldn't possibly be, that pleasure is pleasant? Mill, who undoubtedly thought long and hard about these matters, thought that to call something good and to call it pleasant were just two modes of stating the same fact.[405] The role-property naturalist can make clear sense of the statement: to say that pleasure is good is to say that pleasure plays the value-role, and this might be taken as an advantage of that theory. But the realiser-property naturalist can offer the same answer: the statement is true and informative, but it does not add anything to the *analysis* of 'value'. The naturalist treats the identity statement as parallel to the statement that water is partly made up of hydrogen. This might very well be presenting some news to you, yet it does not predicate some distinct property of water that wasn't in a very real sense "in it" already. In a similar way, we are revealing something of interest, when we say that pleasure "has" value, even if we are not strictly speaking attributing anything new to it.

Essential Value

The search for the nature of value is at least partly guided by a focus on what is *intrinsically* or *finally* valuable: i.e. on what is good for it's own sake as opposed to the extrinsically and instrumentally valuable. Value has often been supposed

[405] In "Utilitarianism" (1993).

to be *intrinsic* to that which has it[406], and the supervenience claim tells us that value is necessitated by the (intrinsic) properties of the value-bearer. Properties that *make* something intrinsically valuable need not be *essential* to that thing, however: a thing may loose its value without thereby loosing its identity.[407] While something else might have been good, if the critical properties where distributed differently – the value role could have been performed by something else - *that role*, or *that property* couldn't have been something else.

Value is not only intrinsic to pleasure, but also *essential* to it. Pleasure could not lack value and remain the property that it is. This is of course trivial if value and pleasure is the same property, but it has some independent plausibility as well. It is not possible to imagine pleasure without the property that makes it valuable.[408] In this sense, the question whether pleasure is valuable is in fact not open. Pleasure is, however, conceivable without some of the properties that make it the *referent* of the term "value", and this is what makes the claim "pleasure is good" informative. It might not have been an intrinsic motivator, for instance, at least according to the theory of pleasure promoted in part 1. Some, as we've seen, have preferred to *define* pleasure as an intrinsic motivator (as "unconditioned reward") and, presumably, would say that pleasure might not have felt the way it does. But that is, probably, a more awkward claim to square with our normal conception of pleasure.[409]

The Argument from Queerness

The notion of a natural property being essentially valuable brings to mind the argument from queerness, which we briefly came across in an earlier chapter. In short, the argument says that value, understood realistically, would have to have

---

[406] Moore (1993). Lemos (1994), See Zimmerman (1989, 2001).

[407] If the value-bearer is not an object, but a state of affairs, however, one could make the argument that everything intrinsic to that state of affairs is essential to it. In that case to have value intrinsically would always be to have it essentially. But we want to assign value to concrete objects as well.

[408] Mendola (1990). Sprigge (2000).

[409] See Berridge (2002). Another alternative is functionalism about experiences: something with this particular motivating function could not feel in any other way. I have my doubts about this view. See my (2003b, under the name Bengtsson), Chalmers (1996).

"to-be-doneness" somehow built into it. Objective values, Mackie wrote, would have to be "objectively prescriptive" and this is metaphysically queer.[410] Since no property we know of is like this, our talk about value is in error. Just what it is that natural properties are supposedly unable to do, is not entirely clear. Even if its proponents take it to demonstrate something important, the argument echoes the complaint that no natural property essentially *commands* the presence of desire the way value should: more or less anything can be, or not be, desired. But if a pro-attitude is somehow *included* in the property candidate, this would seem to cover some ground towards solving this problem. Pleasures, as I argued in part 1, is partly constituted by such an element.[411] If it is objected that this solution provides the wrong value-motivation link, we can reply that this begs the question: no particular motivation/value link is conceptually *required* and we are offering a *revision*, after all. Is it suitably "prescriptive"? Affective experiences, in Mark Johnson's terms, have a form of *authority.*[412] To feel pleasure is to feel something positive going on, and it is usually non-accidentally coupled with behavioural tendencies to maintain it. The link between pleasure and motivation, I believe, means that pleasures *are* kind of "queer", and their experiential nature and their status in motivational matters amply demonstrates this fact. The naturalist hedonist just needs to argue that these facts make pleasures "queer enough".

2.5.3 Hedonism and Explanation

In the last chapter we came across an argument against value realism based on the alleged explanatory irrelevance of evaluative facts. If value facts play no role in the best explanation of our value judgements and beliefs, why should we believe in such facts? If the best explanation of our beliefs need not appeal to their *truth*, they seem to offer no evidence for those beliefs.[413] This complaint is

[410] Mackie (1977, p 38)
[411] As I argued there, this pro-attitude is part of how pleasure *feel* and thus in principle, but seldom in reality, distinct from the functional, dispositional role of being an intrinsic motivator.
[412] Johnson (2001).
[413] Harman.(1977), See 2.3.4. The argument, according to Sturgeon (2005) is that "reference to moral facts appears unnecessary for the *explanation* of our moral observations and beliefs".

particularly devastating for evaluative judgments, since they are often suspected to stem from cognitively unreliable sources like prejudice, emotion, convention, and mere force of habit.[414] The complaint is further fuelled by the persistent disagreements about evaluative matters, making almost any predictor of evaluative judgment better than the nature of the object assessed. Realists faced with this complaint can opt to accommodate apparent disagreements by relativizing the contents of evaluative beliefs and judgments and say that these are relational facts.[415] But they can also say that some of these beliefs are simply mistaken, and that the persistence of the disagreements just means that these mistakes can be very deeply ingrained indeed. The explanatory test would then be a method to weed out the mistaken beliefs about value.

Causal Explanation

Here's how to provide evaluative facts with an explanatory role: The supervenience requirement says that things are good in virtue of "good-making" properties, and we accepted tentatively that these are *natural* properties. Let's grant that these are also the properties in virtue of which we *judge* that things are good. Now let's say that goodness just is this property, or set of properties. Since none would presumably deny that the natural good-making properties have causal powers relevant to explanation, such a reduction would make goodness explanatorily potent.[416] Of course, if value judgments are true in virtue of natural facts it is only to be expected that appeal to value would work in everyday explanations, even if value judgments were not *reducible* to judgments about those facts. They would work as shorthand for, or indicators of, the set of natural properties in virtue of which they are true - or in virtue of which we believe that they are true.[417] If value-properties are non-natural properties and the supervenience relations obtain between distinct properties,

[414] Joyce (2006, 2007).
[415] As Harman does (1977, 1986, 2000), see also Sayre-McCord (1988).
[416] It is noteworthy that Harman's complaint about the untestability of moral claims is conditional on there being no naturalistic reductive account that ties moral facts to natural facts. But see Sturgeon (1985) who thinks no such reduction is necessary.
[417] A similar argument has been made for the explanatory relevance of phenomenal experiences, even if one believe that "qualia" have no such powers in themselves. See Bengtsson (2003b).

value itself would be causally idle, but its presence would imply the presence of natural properties with causal powers.[418] If the natural properties on which value supervene form a naturally congruent set, we have a clear transferred explanatory role at hand. The complaint is that the "goodness" of these facts seems to play no independent part in those explanations, only the good-making characteristics do and it is to this complaint that the naturalist offers the elegant solution: identify good with the "good-makers".[419]

There is a complication with this suggestion, however: It seems that no natural properties reliably make us form the judgment that the bearer of those properties is good. "Goodness", if we take into account all typical first-order evaluative beliefs, doesn't correspond to a causally cogent set of properties. Consequentially, there is no plausible candidate for a naturalistic reduction of value. Gilbert Harman offers a version of this argument, and argues that the only plausible property candidates are *relational* properties. This requires shifting the attention from the things judged to be good to a relation between that thing and some - real or hypothetical, individual or collective - subject. It also requires giving up on one aspect of the objectivity platitude.

The Explanation of Valuation

In a recent paper called "Explaining Value", Harman stated his concern with the *descriptive* side of value, which involves the explanation of why people value what they value, why they have certain moral reactions, such as feelings and motivations to act morally, and why they have the moral opinions they do[420]. This project is clearly one that would profit by joining forces with other scientific disciplines such as social psychology, and Harman is well known for putting such findings to use in philosophical arguments. While philosophers are

[418] They might be directly relevant to the epistemological story, they might be what intuitionists say that we intuit. But even our intuitions are subject of causal forces, why believe that the causal story of their production is not the best explanation of them? What role does the *truth* of intuitions play in an account of our intuitions? This is an interesting and pertinent question that I wont be able to go into detail of here.
[419] This was Jackson's (1998, 2003) point. We might as well be naturalists.
[420] Harman (2000, p 197).

likely to point out that explaining why people value something is theoretically distinct from explaining its value, there still might be significant relation between these two kinds of explanations. In particular, as we have seen, if the best explanation of our valuations need not appeal to their *truth*, we face a conspicuous lack of evidence.

As Joyce points out, every belief has a causal history, and not every such history undermine the confidence we put in the belief.[421] For that story to be undermining value facts must be shown to be irrelevant to the best explanation of value beliefs. It's not sufficient that the explanation doesn't *mention* value facts: Some identity or supervenience relation might hold between items denoted in the explanation and the value properties represented in the belief's content. If such a relation holds, the genealogy might render the belief true after all.[422]

Whether or not moral and evaluative beliefs are subject to observational testing – and we should agree with the sceptic that they are not straightforward empirical beliefs - they are certainly not *uncaused*. We acquire these notions somehow: we learn how to categorize things as good or as bad, and to rank things according to their value in order to facilitate decisions and recommendations. We also seem to revise our beliefs about what's valuable on the basis of (or on exposure to) empirical facts.[423] Even if what we then are doing is learning to attribute a concept we already have some immediate grasp of, there is some psychological arrangement that *constitutes* that grasp. Apart from the obvious intrinsic interest in understanding how this works – is there anything in this emerging explanatory picture that might serve as what these beliefs are *about*?

[421] Joyce (2007).
[422] Joyce (2007, p 184)
[423] There are also other, internal, factors that influence this development . On the moral development of children, See Hauser (2007) and Nichols (2004).

Hedonic Explanations of Value

The hedonist version of this argument is quite straightforward: Pleasure plays a central part in the best explanation of our evaluative beliefs, behaviours and dispositions. In this quite general form, this fact has been the main support for hedonism for most of its history. The argument has normally proceeded from hypotheses about the psychology of desire and motivation[424]: It is its role in *this* explanatory picture that confirms pleasure as being the good.

In its standard form, hedonism is supported by the claim that pleasure is the only intrinsic *motive* for our actions and desires. This view, commonly called "psychological hedonism", is often attributed to Jeremy Bentham[425] and John Stuart Mill. Actions and desires are surely among the things we would expect value to explain, so if this theory is true: so far so good for the naturalist hedonists. But people regularly desire things other than pleasure, and invoke reasons that do not reduce to the expected hedonic effects of their actions. Hedonists are usually aware of this fact, and appeal to revisions of our superficial desires in order to secure two results: to get rid of our apparent non-hedonic desires and to provide a more plausible link between value and desire. Mill, for instance, argued that pleasure is what we *really* desire for its own sake, acknowledging that isolating these desires takes some shearing of what we superficially desire. Mill suggested that we ask about anything that we desire *why* we desire it, and that in the chain of justification that follows, we end up the one non-instrumental object of desire, namely pleasure. Richard Brandt suggested that in order to be value-relevant, desires need to be subjected to (or be capable of surviving) some sort of testing, which he called "cognitive

[424] Bentham (1960), Mill (1993), Sidgwick (1981) on intuitive evidence, See Brandt (1976), Railton (1989), Katz (1986), Gosling (1969).

[425] But note the ambiguities of the following central statement "Nature has placed mankind under the governance of two sovereign masters, *pain* and *pleasure*. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne. They govern us in all we do, in all we say, in all we think: every effort we can make to throw off our subjection, will serve but to demonstrate and confirm it". Bentham (1960).

psychotherapy" – but there is no guarantee that those desires would pick out only pleasures.[426]

Much has been written about Mill's "proof" of hedonism, which moves from the fact that we desire only pleasure to the conclusion that pleasure alone is desirable, and it has been widely criticised, most famously by Moore.[427] Both Mill and Bentham held the value of pleasure to be a basic principle, and that basic principles are not subject of "proof" in any strict sense – it is not a question of *logical* or *deductive* proof. Basic principles are not based on principles that are more secure than they are themselves. Both Bentham and Mill do, however, admit of considerations in favour of basic principles, and it is as such a consideration that desire "proves" desirability. Mill realised that demonstrating a link to motivation does not strictly prove goodness, but noted that nothing else has an equally great claim on weight in this matter. He does suggests that we should turn to the desires of knowledgeable, experienced people, since they can be trusted to have a fair idea of objects that are not present, but it is still ultimately the desire that forms the support, and the judgment just insofar as it is a fair estimate of desire. Mill still holds that the desires of these knowledgeable people track only pleasures, and that their judgment results in a distinction between higher and lower pleasures: the higher, better pleasures are those that a more knowledgeable "critic" would prefer.[428]

Beyond Justification

According to Mill, our inquiry about the good starts with asking what people value and/or desire for its own sake. We find this out by following a chain of *justification:* by asking about anything that we value *why* we value it. We keep

[426] Brandt (1998). Other suggestions have it that the relevant desires are those that are backed up by reasons, or would be held by an "ideal" observer. I agree with Chris Heathwood (2006) that "ideal desires" are the *deus ex machina* of the desire-satisfaction literature. Idealisation only introduced a new notion to be explained. See Rosati (1995).
[427] Mill (1993), see Millgram (2000), Sayre-McCord (2001), Moore 1993).
[428] David Hume offers a similar argument in "A standard of taste", but comes to a different conclusion about the locus of value. See also Sayre-McCord (1997).

asking this question until we cannot find further justification by appeal to anything else. Mill pointed out that the one endpoint we notoriously find in this process is happiness/pleasure. Our intrinsic desire for pleasure is not further justified, and is in need of no further justification – the value of pleasure is self-evident. In this sense, Mill argues, to desire something and to find it pleasant is the same thing, or: the expressions refer to the same psychological fact.

Mill acknowledges that some people might genuinely believe that things other than pleasure are valuable for their own sake, but there is always some explanation of this: there is something amiss with such agents or the agent somehow misrepresents his own internal state as being a desire for some external good, when it is actually for some part of happiness. The former point is substantiated by the fact that non-hedonic intrinsic desires are not universally shared. Since the theoretical project is concerned with subsuming apparent goods under as general principles as possible, we should disregard such anomalies.

A distinguishing mark of this argument is that the explanation it appeals to involves features of conscious reasoning and justification. I doubt that such an appeal will be supportive of the hedonic case, for two, very much related, reasons: there is no reason to expect that our desires, no matter how we interrogate them, have pleasure as their only object. And, more importantly, this mode of justification might very well be misleading as to what *actually* determines our desires and evaluative judgements. The main difference between Mill's argument and ours is that we follow a chain of *explanation*, not justification. It is as what *causes* us to value certain things, rather than as what we take to be the reason to value them that pleasure forms a suitable end-point of the argument.[429] The best reason for this shift of attention is the lack of a suitable universality to the justification people offer, and the fact that we clearly

[429] There are, of course explanations of *why* we experience pleasure at certain points, the explanatory chain does not just suddenly come to a stop. There are, for instance, evolutionary reasons why we have a system for evaluation, at all, and why it behaves as it does. I merely suggest that our best candidate for the role of *value* exists at this point in the explanation. In addition, there are certain problems with the explanatory force of evolutionary psychology which it would take us to far treat here.

can be mistaken about the foundations of our evaluative beliefs, seeing how easily manipulated our evaluative system is. Mills strategy would have some measure of credulity if there were such a thing as a common end point for justification. But even then, if that object did not figure in the best *explanation* of our evaluations, we should be suspicious of the force of the theory.

Vindication and Undermining

Hedonism implies that many of our attributions of value are false. This is not merely the claim that other *theorists* have been mistaken about the nature of value; it is the claim that most everyday applications of the term, even those sanctioned by practice, systematized through use, and survivors of serious introspective questioning, are false. How can a theory like this be a *vindicating* version of naturalism?

It is not hard to imagine someone having fundamentally false beliefs about what's good. We can even conceive of worlds entirely populated by people with false beliefs about value: we only need some way to determine the reference of value in a way that does not fluctuate between agents and worlds. But can it be consistently claimed that *our* beliefs are systematically false, and still hold on to a realistic theory of value? Hedonism would seem to violate the general premise that value cannot be radically different from what we think it is. But consider this: Most of us have no problem with the idea that people are, or have been, mistaken about the nature of matter, of the human mind, of time, the universe etc., so why should it bother us that the same applies to value? A theory just has to make sure that there is some continuity between the content of the belief we aim to correct and replace, and the content of the account offered: the theory needs to target much of the same *problems* as the theory it aims to replace did. Scientific developments can be said to continuously change the subject, but we are quite willing to do so because of the insufficient grasp we had of the subject to begin with. Naturalistic hedonism depends on granting this same status to the problem of value.

Explaining and Explaining Away: Dislodging Intuitions

The fact that meta-ethics is in a state of perpetual and fundamental disagreement practically guarantees that, as David Sobel puts it, any theory of intrinsic goodness is sure to rub some of our intuitions the wrong way.[430] While regrettable, this also means that prima-facie counter-intuitiveness is not that great a theoretical burden. Hedonists in particular should hold on to this fact, seeing how hedonism is often rejected on intuitive grounds.[431] Sobel points out that when a theory clashes directly with central intuitions, it can regain plausibility in two ways: 1) By *clarifying* our intuitions - When we properly understand what the intuition is actually about, we find that it is compatible with the theory. 2) By explaining the intuition *away* - by telling a story that undermines its credibility.[432]

Both these strategies have been present in the hedonist literature at least since Mill, and very probably from the very start.[433] The scientific research into matters of evaluation - particularly concerning the role and function of emotions - is an excellent tool for dislodging intuitions that appear to be at odds with our theory.[434] To some extent such research can be used to *correct*, to some extent even to *replace,* some of the intuitions that govern meta-ethics by offering more reliable methods. At the very least, we should be prepared to change our mind about the relevance of our intuitions as a result of an investigation of where these intuitions come from. Intuitions may still be the ultimate justificatory endpoints even if we realise that not *any* intuitions will do as such an endpoint. They are acceptable as endpoints only when we can see no

---

[430] Sobel (2002).
[431] See also Mendola (2006), Moore's reply to Mill and Sidgwick (1993).
[432] Sobel (2002) p 244, other such treatments of intuitions in Railton (1989) and Brandt (1976).. Sobel finds this argument flawed, since it means that we cannot introspect the truth of hedonism.
[433] Epicuros points out that we regularly project value on things that merely tend to cause it. Or rather, on things that we *believe* cause it, seeing how we also frequently misjudge what makes us happy.
[434] See Much and Klauer (2003) Zajonc (1980), Schwartz and Clore, (1983, 2003) Kahneman, Diener, Schwartz (eds) (1999)) , Berridge (2002, 2003, 2004), Kringelbach (2009) and Haidt (2001).

other way of providing support for them, and when they are not undermined by their explanation.[435]

2.5.4 Hedonic Psychology and Psychological Hedonism

In his 1986 dissertation Leonard Katz argued that hedonism should be based on a scientific account of the role pleasure plays in human nature, in action and in evaluation more generally.[436] The nature and function of pleasure reveals it as having the sort of status that a candidate for being what makes life good must have. It *supports* pleasure's "presumption of being intrinsically good", while giving no support for the value of other objects. Indeed, the hold other objects have on us is regularly revealed to be dependent on hedonic processes. When non-hedonic values do exhibit some measure of independence, the hedonist can chalk this up to habit, to misattribution based on hedonic experience, or to some extrinsic system of norms and sanctions. The presumption of pleasure to be intrinsically good is partly substantiated by its initial plausibility as a substantive good, but also by its function in evaluation more generally.

Searching for empirical support for a theory of value lead us to the affective and cognitive sciences, and to social and behavioural psychology. In these disciplines, there is an increasing support for a central role for emotion in evaluative processes.[437] To some extent what's found in recent scientific studies have been known - or more accurately: guessed at - throughout the history of philosophy.[438] This is hardly surprising seeing how psychological processes are,

[435] See Joyce, (2006) and Cullity (2006).
[436] He also pointed out that the separation of ethics from psychology and metaphysics and the separation between questions about the *value* of pleasure and its role in human nature has "barred the way to their understanding just as surely as their confusion did in centuries past" (Katz (1986) p 44).
[437] See for instance Damasio (1994), Epstein (1994), Slovic et al (2007) Haidt 2001 , Prinz (2006) .
[438] This is, by and large, the *sentimentalist* approach, and it has proponents at least as far back as in the works of David Hume. See Hume (1978), D'Arms and Jacobson (2000), Subjectivism in Wiggins (1998). See Prinz (2006) on the relation between sentimentalism and intuitionism. The complaint about philosophical guesswork about motivation was furthered by Stich and Doris (2006).

in a way, supremely accessible to us – via introspection and observation of others. But the more precise nature and function of the processes of evaluation may not always, and not entirely, be introspectively accessible. We are liable too certain cognitive mistakes, when it comes to evaluation: we tend to exaggerate the importance of certain factors and downplay others, and in general to rely on intuitions whose function is not primarily to facilitate value-theoretical lucidity, but chances for survival and competitive edge. It is in this regard that empirical research can facilitate progress towards solving certain problems in value theory.

While it is clear that pleasure plays a central role in the process of evaluation, what role this is, in particular with respect to motivation and deliberation, is still very much under investigation.[439] The various sciences concerned with motivation reveal a number of ways in which hedonic qualities and processes account for differences in evaluation and for differences in other areas of cognitive processing as well.[440] One important point is that the role played by pleasure is not always that of a *goal* of our actions. In fact, its function is not primarily facilitated by pleasure being the *object* of motivational states[441]: our honest intrinsic desires might very well be directed at things other than pleasure. If the plausibility of hedonism depended on *that* relation to desire, hedonism would be undermined by the facts of the matter. [442]

The fact that we intrinsically desire other things than pleasure means that the intimate connection between value and desire needs to be of another sort, if hedonism is to get any support from it. As we saw in part 1, one such proposal involves recognising a desire element in constituting the nature of pleasure. The

[439] See for instance Slovic et al (2007) Kringelbach (2009), On evaluative conditioning, see Jones, Fazio and Olson (2009) – "People are commonly observed making mistaken attributions about their psychological experiences, and this seems particularly true of affectice or evaluative experiences".
[440] See Richard Davidson on how mood influence cognitive functioning, whereas feelings influence action more directly ((1994) Berridges survey "motivational concepts in behavioural neuroscience" (2004).
[441] Moore (1993) was onto this, and stated outright that there is probably some "universal connection" between pleasure and desire, but he sensibly believed that it was just not *this* one.
[442] See Heathwood (2006, 2007), Sidgwick (1981), Feldman (1997a, 2004). See also part 1.

other move involves the explanation of intrinsic, but external, desires.[443] Happiness, as Richard Brandt points out, is the natural source of all our desires and aversions.[444] When we associate pleasure or happiness with a certain experience or state of affairs, that experience or state of affairs becomes the object of desire. The desires developed by an individual can be at least partly explained as "tracing a path towards the experience of happiness", even though they themselves often do not aim for happiness. It's just a fact about human nature that we come to like for themselves things that reliably lead to other things that we already like. The process by which pleasure influence desire and evaluation more generally is most likely even more variable than Brandt suggests; for one thing: association learning is not merely conditioning: We can come to desire things via their association with pleasure, even if we do not even implicitly believe them to be the *cause* of that pleasure, and we can form such desires whether the association is noticed or not.[445]

Pleasure and the two systems of evaluation

In the psychological literature, it has become increasingly common to talk about two systems for evaluation.[446] One system is broadly "emotional" – a matter of reflex-like, unconsidered responses. The other is broadly "rational", associated with higher cognitive processing and justification. The more realistic picture is of course more complicated, but there is something profoundly relevant in this distinction. Epstein develops this notion of evaluation as a dual processing system, and qualifies it so that the emotional system is experiential and affective in nature. He holds that this system is responsible for most of our actions and attitudes, whereas the "rational" system is mostly concerned with explaining and justifying our behaviour to ourselves and to others. The subject

[443] Harman (2000) points out that the plausibility of utilitarianism depends on the place of happiness in the explanation of our moral judgments. It's not that our moral reaction is to calculate the happiness-maximising action. It is rather that our moral reactions are *sensitive* to the way pleasure and pain is caused by the action under assessment.
[444]Brandt (1998), p 100, Also Railton (1989).
[445] Via so called "Affective priming". See Slovic et al (2007). Schwarz and Clore 2003, de Houwer (2007) on evaluative conditioning.
[446] Damasio (1994), Haidt (2001), Greene, (2001), with Haidt (2002) , Epstein (1994), Ledoux (1996).

who is in the affective state might be unaware of this fact, even as the experience motivates actions and regulates thought - typically to maintain that state, if it is pleasant, and to avoid it, if it is unpleasant.[447]

While most people recognise that pleasure and pain are strong motivators, and hold pleasure and the absence of pain to be things worth striving for, everyday life doesn't necessarily involve awareness of how hedonic states and processes influences our dispositions and judgments, be it directly or via long-term programming effects. This is an implication of what Paul Slovic calls the *affect heuristic*. Rather than weighing the pros and cons of particular actions in particular situations, something that requires retrieval of memory of a number of relevant examples, we use an overall, readily available affective impression. The need is particularly pressing in situations when the situation calls for a complex judgement, and the mental resources available are limited. Affective states can thus serve like a mental "short-cut", a "heuristic".[448]

Calling it a "short-cut" should not mislead us into thinking that our knowledge of what is good is somehow prior to, and established independently of, this affective labelling. It's not as if we have independent values that we then come to *associate* with hedonic experiences. Things *arrive* as pleasant, this is how we get to know what is important, and what we later use the secondary system to deliberate *about*. The secondary system is in place for long-term planning, but even this is largely done by the first system: Affect is called a heuristic, after all, because of its pragmatic properties.

It is important to realise that while driven by pleasure and pain, the organisation of our motivational system is in no way guaranteed to maximize pleasure and minimize pain.[449] But neither, as is well known from the "paradox

[447] Epstein (1994, p 716).
[448] Slovic et al (2006, p 1336)
[449] Slovic et al (2007): …if it was always optional to follow our affective and experiential instincts, there would have been no need for the rational/analytic system of thinking to have evolved and become so prominent in human affairs.(p1347).

of happiness", is striving towards pleasure always the best way to attain it.[450] We are flawed as predictors of what will make us happy, and liable to systematic mistakes in calculating outcomes.

None of these findings implies that "higher thought" is somehow *opposed* or *counteracts* the hedonic system. Higher cognitive systems, largely associated with activity in the prefrontal cortex, *modulate* hedonic experiences and its influence on motivation, and are in turn modulated by it. Nevertheless, their separation means that conflicts can and do occur. Direct opposition occurs between what we think we should do, or what we think is *good*, and what is most hedonically tempting. There are certainly cases when non-hedonic reasons promote hedonic results, and the hedonic temptation would be disastrous. But for the most part, these faculties work together, and we are probably best of when they do. Reason on its own, as Hume argued, is not sufficient to account for motivation: we need some valenced outcomes to begin with, which can then be handled by reason in order to bring it about, maximize or distribute them, as something to indulge in or to strategically withdraw from. The hedonist needs only to claim that the *origin* of this process is found in the first, emotional/hedonic system.

Pleasure and Misattribution

Pleasures are frequent objects of intrinsic desires, but, as we've seen, they can themselves be (constituents of) intentional states that take external objects of their own. Not surprisingly, then, we tend to assign value to the object of that experience, or, if the experience as such has no object, to the most salient object around. We tend to value things that we get pleasure from, or believe to get pleasure from, and disvalue things that cause us suffering, or that we believe to have caused these things.[451] Affective states are used as *evidence* for the value of the object under assessment. This was one of the findings of an influential study

[450] Sidgwick in ch. 4 of the Methods of Ethics (1981). See Kahneman (1999), Nettle (2005), Layard (2005).

[451] The process sometime goes under the term "evaluative conditioning", which focuses on the conditioning of *attitudes* rather than on behaviours, as in classic Pavlovian conditioning .See de Houwer (2001), (2007). Jones, Fazio and Olson (2009).

conducted by Schwarz and Clore in 1983.[452] In this study, people were asked about how satisfied they were with their life in general (i.e. not in that particular moment). The question was asked under conditions where external factors (the weather) had a non-noticed impact on their mood. The authors found that evaluative judgments often involve people implicitly asking themselves "How do I feel about this?" and in doing so, read their current feelings as a response to the object of judgment, whether the cause of that feeling is relevant to the value of that object or not. Importantly, this effect does not occur, or not as strongly, if the informational value of the feeling is somehow discredited, as when we correctly identify the source of the feeling and find it irrelevant to the object under assessment. If the subjects interviewed were made to pay attention to the weather, the emotion/valuation correlation was severely weakened, but only when the impact was a *negative* one.

Clore and Huntsinger write that according to the affect-as-information hypothesis, affect *assigns* value to whatever seems to be causing it. While theorists often assume that people's attitudes and judgements reflect information about the object of judgement. "…people's evaluation also reflects information from their own affective reactions."[453] It is noteworthy that people are more likely to revise their judgment when the feeling and the attribution are negative than when it is positive: it's more important to identify the source of bad feelings than of good ones. In the first case, something needs to be done, and in the second we are happy to stay were we are.

Our evaluations are not entirely as fickle as our feelings: evaluative judgments can attain independence from the process that instigates them. The hedonic process can *establish* the value of an object to make it cognitively relatively independent of that process. This is familiar from classic conditioning, where conditioned rewards can remain operational, even when the unconditioned reward (pleasure) is no longer forthcoming, at least for a while. For normal subjects, and if there are no secondary reinforcers in place, a hedonic reversal

[452] (1983), revisited (2003).
[453] Clore and Huntsinger (2007), (p393)). See also Williams and Bargh (2008).

will eventually dislodge the value of an earlier perceived value. Because most values are socially, culturally, sanctioned, there remain strong reasons not to go entirely with the feeling - there are inherent obstacles in this system, and benefits of displaying consistency and the ability to plan ahead.

Certain external values can be so hardly ingrained that we would not abandon them, even when they cause us suffering. Addictions such as gambling might be like this, as might obsessions more generally.[454] Some evaluative dispositions seem to be hardwired, but most are wired to concur with what's pleasant or unpleasant.[455] Such plasticity is needed in an environment as flexible as ours, an environment that depends to a significant degree on other, equally flexible, beings. The dependence is not always directly noticeable: A sudden hedonic disappointment, or reversal, would not necessarily change our mind about the value of an object – but note that strong unpleasurable conditions like stomach-sickness can put one off a previously favoured food-stuff forever. *Persistent* hedonic disappointment or reversal tends to influence decisions and evaluations in normal subjects in precisely this way.[456]

Evaluation often occurs in *social* contexts, and this requires finding some independent, non-psychological common ground for their justification. Justification is also, largely, *informational*. We search a common, objective ground, independent of psychological events, because it is by paying attention to such external object that the hedonic potential is realised. It might be a good thing that we don't realise the mere instrumental value of the things we value. This is partly due to the fact that extrinsic motivation tends to undermine intrinsic interest[457] – if we realised that the value of some activity we're engaged in is extrinsic to that activity itself, we might stop being into it enough to

[454] See Berridge (2002, 2004), Kringelbach 2009).
[455] Normally, hedonic processes are involved in learning processes that are then hard to change, once the association has been firmly established. Evaluative Conditioning is not as sensitive to extinction as classic behavioural conditioning.(de Houwer (2001) – "empirical evidence suggest the majority of likes and dislikes are learnt and not innate).
[456] It is a *big* problem when this does not work, as in overeating. See Kringelbach (2009).
[457] Lepper, Greene, Nisbett (1973).

ensure its valuable result. Taking an intrinsic interest in extrinsic values has a tendency to generate actual intrinsic value.

2.5.5 Hedonism and the Experience of value

Reductionist versions of naturalism often claim that goodness is identical to a natural property that might be otherwise identifiable, i.e. one that we can also know under some other description. This dissociation has been held to be objectionable, and might indeed be the psychological foundation for the Open Question Argument.[458] The argument depends on our having some form of non-inferential knowledge about value to begin with, and one influential suggestion has it that our access to value comes via a faculty of *intuition*[459]: while our knowledge of the natural properties of an object or a state of affairs is acquired via the normal senses, or derived from sensory information, their value is known via intuition.

Hedonists need not deny this claim. Pleasures are essentially evaluative experiences and as such just *are* this experiential, non-inferred "knowledge" of value; it *is* the "intuition" of value. Pleasures are *value experiences,* or, in more familiar terms: *pleasure feels good.*[460] As opposed to the notion of *being* good, it is quite clear what "feeling good" means, we are immediately acquainted with it. The hedonist, then, should make the case that this is the fundamental value-phenomenon. Understanding pleasure as the experience of value means that we have a direct, evaluative experience of a motivational state, demonstrably integral to the processes surrounding most things associated with the term "value" in everyday parlance.

Joseph Mendola argues that the normative properties of value and disvalue belong essentially to the phenomenal experience of pleasure and pain.[461] A

[458] See discussion in Sturgeon (2003).
[459] Developed by Ross (2002), Moore (1993),Sidgwick (1981).
[460] C.I. Lewis (1947) called it "a value-quality found in the directly experienced".See Bengtsson 2004, Aydede (2000), Mendola (1990, 2006), Sprigge (2000), Epicuros. Prinz (2006) points out that sentimentalism and intuitionism used to be considered to be in conflict, but can be squared with each other if emotion is recognised as a form of intuition.
[461] Mendola (1990).

complete description of pleasure requires what he calls a *committing mention* of objective value, as distinct from the non-committing mention of properties in the context of propositional attitudes: belief in witches is compatible with there being no witches, but the experience of value cannot exist without there being value. As he points out: "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".[462]

While experiences are often quite complex, this evaluative quality, Mendola argues, is a single qualitative aspect that can change independently of other constituents of the experience. On this understanding, while we can always say that one experience is better than another, it's hard to make sense of the statement that one pleasure is twice as pleasant/good as another. I agree that there is clear sense to be made of this notion of "betterness", but in contrast to Mendola, (and in agreement with C.I. Lewis) I believe that even though the experience of value is this simple phenomenal quality, there are still many *different* ways in which one may have more or less of it: It can be more intense, more pervasive, take up more of consciousness, etc. Identifying value with pleasantness, doesn't tell which of these variables "really matters".

The trick here is to treat the evidential weight of these affective intuitions not by their alleged *content* but by their *function.*[463] Evaluative intuitions thus understood are quick, action-guiding and attention-shifting mental states, and this need not have much to do with their content. Content is something we come to *associate* with them. As mentioned, Mackie[464] thought that the phenomenology of value presents it as an objective feature of the object of the experience, and since there is no corresponding property, talk about value is in error. But it is actually quite hard to make uncontroversial sense of the content of phenomenological experiences; what we *do* associate with them concerns their function and some non-conceptual grasp of how they feel. The direct

[462] Mendola (1990, p 702).
[463] As argued by Prinz (2006).
[464] See Smith (1994), McDowell (1988).

"epistemological access" granted by the experience of value is not to a proposition: it is not an experience with the content *that something is valuable*. To say that it is the experience of value is rather just a name of how it feels, not about what it tells. It is comparable with perceptual experiences insofar that the experience of red is not, arguably, an experience with the content "x is red", but, at best, provides *evidence* for x being red. There is a superficially similar relation between the experience of value and value judgments and attitudes, but there are important differences as well. What causes the experience of value varies with interests, with context, with cognitive and emotional background and it varies with the previous attitudes of the subject. The case for identifying pleasure with goodness depends on its role in shaping these judgements and attitudes that are existentially independent of the occurrence of affect.[465]

Pleasure and the appraisal-theory of emotion

The view that emotions are essential to understanding value and evaluation, both functionally, by what they do, and phenomenologically, by how they feel, has a long history, and has received additional support in recent years.[466] While giving this matter due attention would require a more extensive treatment than I'm giving it here, I will risk making some brief remarks about the importance of this research and these theories the hedonistic project.

According to the *appraisal theory,* emotions are individuated by how they appraise changes in the internal and external environment.[467] Emotions are interpretation of detected change. Is it good or bad, is someone responsible, is it familiar, urgent, under control? Emotions are rich experiences that can be experienced as complexes of aspects, but they can also be described as occupying a single point in a multidimensional space; distinguishing these aspects is not a prerequisite to have the emotion. Emotions are not to be understood as *conclusions* from prior cognitive interpretations; while they *can* be reactions to

[465] We categorize things from their influence on emotion as well as on their own characteristics (See Forgas, 2003).
[466] Damasio (1994), Slovic et al (2007), Prinz (2006).
[467] See Ellsworth and Scherer (2003).

cognitive processes, they often are themselves prior to, or part of, cognitive processing.[468]

Now, the appraisal that distinguishes emotions from other mental states is whether what is going on is experienced as something *good or bad*. This evaluative component is intrinsic pleasantness, or *affect*. This is in contrast to the James-Lange tradition, which took emotions to be perceptions of internal changes of the body; things like change in pulse-rate and breathing. It did not recognise that virtually *any* perception of bodily changes can be experienced as good or bad, and that this is a distinct part of the emotional experience, consisting in the feeling of *affect*.[469]

Ellsworth and Scherer point out that in order to survive, it doesn't suffice that an organism understands its situation: it has to be motivated to do something about it. While some species have solved this problem by having fixed responses to stimuli, the emotion system affords a more flexible alternative. Emotions, they write, "imply action tendencies without complete rigidity".[470] While affective reactions might come with a default setting, they are sensitive to changes and influenced by the motivational state of the agent, its preferences and goals. Intrinsic pleasantness provides *general* guidance - approach or withdrawal - but we need more information to pick a specific goal and response, and for this we have a more complex system of motivational appraisals. The appraisal model recognises that motivational relevance and the goal conduciveness of a stimulus is distinct and can even be opposed to pleasantness, as when a pleasant stimulus distracts from a more important task. Still the appraisal model supports the basic explanatory role of hedonic states like pleasure and displeasure, and further shows how other cognitive processes modulate this role.

---

[468] See Zajonc (1980): Feeling and thinking: preferences need no inferences. Feelings often precede thinking.

[469] One reason is that James couldn't find any dedicated *organ* or *faculty* for affect. Since then, dedicated parts of the brain, neural pathways and neurotransmitters have been identified, responsible for affect. See Damasio (1994), Berridge (2002, 2004), Davidson (1994), Katz (2008).

[470] Ellsworth and Scherer (2003, p 572).

2.5.6 Response-dependency and Pleasure

According to one influential theory, the way to understand the property and/or concept of value is via the notion of *response-dependency.*[471] To be good, according to the schema for this view, is to stand in a certain relation to a subjective response in an agent. Goodness may then be understood as the property *that* stands in this relation to the response, or as the second-order property *to* stand in such a relation. The schema is applicable to other properties: Disgusting things are those that cause disgust in certain observers under certain circumstances. Indeed, since disgust is essentially grounded in the response in this relation, we can easily vary the other entities in the relation: It makes sense to say that something is disgusting to *me*, or *right now*, but we can also apply it to what is disgusting to normal observers under normal circumstances, or to observers equipped with a certain sensitivity etc.

Another variable is the type of relation referred to. We want to say that things are disgusting that *would* cause disgust if encountered: to be disgusting is a *dispositional* property. This means that the *property* of being disgusting is not response-dependent, since it does not depend on anyone ever responding to it: dispositions are usually ontologically independent of the conditions for their manifestation. Only the *concept* would then be response-dependent.

The relationship between property and response may also be a *normative/epistemic* one. Moore famously argued against what he took to be Mill's view that the desirable is not what we *do* desire[472], but what we *ought to* desire.[473] Insofar as appealing to the desires of an enlightened agent is a plausible qualification, it is because such agents are more likely to get their desires *right.* This can be understood in different ways: perhaps desiring what is good is like fearing what is dangerous: it's something the good *merits*. But Equipped with a pragmatic, evolutionary take on psychological sensitivities, we might say that the disgusting is what we *should* react to with disgust, because it

[471] One that received some sophisticated treatments in the 80's and 90's: McDowell (1988), Wiggins (1998), Johnson (2001), Railton (2003), Goldman (1987), Lewis (1989).
[472] Or what *causes* the desire, note that what causes a desire need not be the object of that desire.
[473] Taking a normative relation to be the point is familiar from the "buck-passing account" of value, related to this schema. See for instance Scanlon (1998). But see also Crisp (2005).

is detrimental to our health. Disgust, according to this view, has a *target* and might be seen as appropriate or inappropriate, or even true or false. For a number of responses, emotional and otherwise, we can conceive of them as adaptive responses to objective features of the world that are metaphysically independent of that response. Fear can be understood as an adaptive response to the dangerous. Even though the bi-conditional holds: being dangerous is to merit fear, this does not mean that being dangerous is a response-dependent property, nor a response-dependent concept. Or, rather if it *is*, the response that defines it is not fear, but being hurt, or dead. The response-dependent property/concept associated with fear is rather being *scary*.

If value is a response-dependent property/concept, we need to fill out the schema with the relevant response, subject and relation. There seem to be almost universal agreement that the relevant response is a pro-attitude of some sort, whether that is understood as desire, admiration, or as any response that has an evaluative function or content. If this sounds viciously circular[474], keep in mind that while we have difficulties accounting for what *value* is, *evaluation* seems fairly straightforward. Not in the sense that their content is obvious, but in the sense that we are *familiar* with them. Like the experience of colour is epistemologically prior to colour itself, the evaluating emotional responses are epistemologically prior to value. This is why we turn to evaluation to identify our subject matter.[475] The things we respond to in this way seem to have nothing *else* in common that would serve as what these responses *detect,* nor have we any reason to believe that more enlightened agents would respond to some naturally distinguished group of objects in this way. If the distinctive feature of the relation between response and object is thus found on the response side, we can easily understand how a range of semantically related value-concepts might arise. Something could be good for me or for the group

[474] See Wiggins, who agree that this is circular *if offered as a definition or analysis*, but that it is true and informative nevertheless.
[475] Even if we recognise the "authority" (Johnson (2001), McDowell (1988) with which this response might seem to *reveal* something about the object.

I'm in, or for a subject defined in terms of full information and absence of cognitive infirmities.[476]

We should, I believe, resist the tendency to pinpoint value by focusing on the object, because the objects have nothing suitable in common. We should resist focusing on the relation, because a normative relation leaves us with an unaccounted for, mysterious notion. And we should resist focusing on the subject because no particular feature of it is likely to distinguish value from any other response-dependent notion. We already have an irreducible evaluative element in the "analysis", and it belongs to the response. What makes the response evaluative is what makes an experience evaluative: i.e. pleasure.

Where in the world is value?

Most response-dependency theorists place value in the object of the response, rather than in the response itself.[477] The rationale for this is basically the argument for naive realism: the properties that we perceive are not in the head, but in the object, where the act of perception puts them. In reply to the objection that different objects may give rise to the relevant response in different possible worlds, resulting in relativism, the suggestion is to *rigidify* from the response to the object: the good is what *actually* stands in the preferred relation to the response, not what people in other hypothetical situations would respond to.[478]

We considered a rigidifying relation between the concept of value and whatever deserves the name in chapter 2.3, and mentioned some considerations for and against it. One of the considerations had to do with retaining an intimate link to motivation, something that might belong to the functional characteristic essentially, but only contingently to the property picked out by the function.

[476] See Brandt (1976), Hume (1874-5).
[477] McDowell (1988), Railton (2003). But see Heathwood's (2006) and Feldman's (1997, 2004) versions of hedonism.
[478] There are corresponding versions of relativism that use this strategy, see Lyons, Sturgeon (1994), Sayre-McCord (1991).

This seems to speak against rigidifying from the response, because it is the response and not the object that is motivational. I believe that a rigid relation holds between the conceptual description and a response-like mental state, and not between that description and the object of that response. We should, I believe, say that we share a concept with anyone that has the relevant response, what we regularly disagree on is the cause for the response and *that*, I am here to claim, is largely a contingent matter (still one worthy of consideration, of course).

The naturalist hedonist claims that value and pleasure is the same thing. But what should we say if motivation and pleasure come apart? Would we want to say that pleasure still is what 'value' refers to? I say yes, but it is a close call. The reason is that pleasure, while in principle dissociable from functional motivation would still *feel good*. If we imagine a world where the experience of green or, nightmareishly, pain, is what drives motivation in the way pleasure actually does, we should feel sorry for the inhabitants[479], because pleasure would still be what's good. Our conceptual liberalism, treating conceptual features as negotiable, and even as matters of theoretical taste, allows us to say that they have *a* concept of good, but that it differs in some significant respects from ours. This is preferable to saying that they have no concept of good; or are constantly mistaken about how to apply it.

There are several reasons to prefer hedonism to object-focused "response dependence" accounts. It affords a direct epistemological access to value that need not posit any mysterious external property. The ontology is more convenient, pleasure is a more natural property than the property to cause pleasure. We want value to be both objective and sensitive to individual differences, and the hedonic account gives us that. Basically, my point is that the response-dependency account is plausible because a particular form of

[479] They, presumably, wouldn't, at least not if qualia are epiphenomenal (See Jackson (1974), Chalmers (1996), Bengtsson (2003b).

response makes sense of value, and this response is most plausibly given a hedonic interpretation.

2.5.8 Criticisms

So far, this has been an almost exclusively positive account, not facing up to the many objections that have been lodged towards hedonism throughout its history. I've wanted to be quite clear about the kind of claim the theory is making, and what kind of foundations it rests upon so that we won't have to encounter objections based on misunderstandings. I believe that the version of hedonism presented, with the explanatory devices laid out, is capable of answering, occasionally to avoid, the classic objections. The objections come in roughly two forms: first, objections to naturalism in general, and to the empirical approach. I hope I have met those objections in previous chapters, and I will say something more about them in the next section. Then there are objections towards the *substantial* theory, the view that only pleasures are good. It is to those I now turn.

The problem of other values

There are basically two objections against substantial hedonism: Pleasure is not the only good, and not all pleasures are good. If this is not immediately obvious, the literature provides thought-experiments to coax this intuition out of you. Nozick's famous *experience machine* argument is used as a criticism of all mental state theories. Imagine a machine that you could hook up to, which you were convinced would give you the experiences you would want - including experiences *of* what you want (there is no restriction on the kind of desires you have). If experiences are all that matters, the life in the experience machine could be made as good as you'd like. Now, would you hook up to this machine for life?[480] Would it be *good* if you did? The intended reaction is clear: of course not. Not only does Nozick predict that no sensible person would agree to be

[480] Nozick (1974). If you want to avoid the science fiction, a very similar argument asks if being successfully deceived is as good as the real thing. Griffin (1986).

connected to the machine for life, authenticity, he claims, is important whether you care for it or not.

The hedonist has a number of replies to offer[481]: First: Nozick is right; we care about other things than experiences. The hedonist is only bound to accept that the life in the experience machine would be *good*. If this still sounds counterintuitive, we turn to the two strategies, look at what these intuitions actually *say*, and see if they can be explained away. First, we have no moral obligation to hook up to the machine. In it, we would presumably not *do* good, we would not attend to our projects and interest. While the hedonist argues that these have no actual intrinsic value, they might still be things that are *important* to us in the sense that we are *interested* in them. As we have seen, the hedonist case depends on getting the right kind of connection between our interests, desires etc. and value, and it turns out that hedonism is, in fact, not undermined by our sincere desires having objects other than pleasure.

We are quite sensibly biased against things like deception, inauthencity, life-time contracts and disconnection from our normal surroundings, these are useful dispositions to have, but these reactions track strategic facts about our psychology, not about the nature of value.

"Bad" Pleasures

A claim is often raised that some pleasures, notably *malicious* ones, have no, or even have *negative* intrinsic value.[482] How should we account for this claim/conviction? As mentioned, there are instrumental reasons to avoid some hedonic temptations, and there might even be reasons to be disposed to shun some pleasures *intrinsically*, as it were, to avoid temptation. It seems to be a fact that some activities are notorious in their tendency to create new opportunities for pleasures. Interestingly, these are things like friendship, curiosity driven learning, sex, art etc., i.e. things that are often judged to be good for their own

[481] Some of which are already present in Katz (1986), Railton (1989), Kawall (1999), Silverstein 2000. See also Sobel(2002).
[482] Ross (2002), Lemos (1994), Feldman (2004), Goldstein (1989), Zimmerman (1980).

sakes.[483] There are also things that are notoriously damaging, among which are enmity, addiction, jealousy etc., even though they might offer some pleasures to begin with, they are likely to turn sour. This would seem to explain why we take some pleasures to be bad. There is also the fact that malicious pleasures seem *morally* objectionable: but this does not undermine the fact that they are good. In fact, it would seem that it is *because* some bad person is getting something good, that he doesn't deserve, that we are dissatisfied with the state of affairs.[484] The hedonist, as we've seen, points out that we *project* value to objects, and there is nothing strange in a state involving pleasure, even our own, provoking negative emotions, and vice versa. I take these cases to support the value-role of pleasure, even if the emotional state itself misplaces this value elsewhere.

The Repugnant Conclusion

The theory proposed is a theory about value, not about morality. Nevertheless, it has features that might enlighten certain problems in moral philosophy, in particular a problem for *utilitarianism*. In his 1984 book "Reasons and Persons", Derek Parfit points out that a theory that says that more pleasure is always better is committed to the following claim: Consider two worlds: In the first is a world were everybody enjoys life to a great extent, and no one suffers. In the second, people have lives that are barely worth living, their hedonic state is such that they only just pass that level. Now, the question which of these worlds is better, if hedonic utilitarianism is accepted, seems to come down to only this: the relative size of the population. If the population in the second world is just made big enough, it will contain more pleasure than the first world, and thus, according to the theory, be better. This is the "repugnant conclusion".

A number of solutions have been proposed to this, some want to say that the highest *average* level of pleasure is what matters, others that the *quality* of pleasures need to be worked into the calculation of the value of a world, but

[483] See surveys in for instance Layard (2005) and Nettle (2005).
[484] See for instance Silverstein (2000),

there are problems with these solutions to. If only average level matters, a world with only one very happy person would be better than a large population of just slightly less happy ones. The quality view offers similar problems.

The solution, or rather "diagnosis" that the version of hedonism I favour offers is this: there is *more* of the good in the repugnant scenario, so it is somehow better. But the world were people are happier, but fewer, is more pleasant in a *distinct sense*. Just like a world can be "more red" by having more red things in it, or by being entirely red, or by having more intense red things in it: there is not *one* dimension of "redness", and they cannot be worked into a formulaic exchange rate. Value, if understood as pleasure, is variable in at least as many ways. Hedonism, in this basic outline, doesn't say and *shouldn't* say, which of these dimensions *matters*. There might not even be a *fact* about which matters. Pleasure and value is the same thing, but that doesn't settle what to do with it, whether to maximize it, how to maximize it, whether to distribute it equally or according to desert.

3. The good enough

This book has pursued a reductionist, naturalistic project. I've tried to show that a naturalistic meta-ethical framework, and an explanatory, empirically informed approach to value, comes out in support of hedonism. It is with pleasure, I believe, that value *enters* into the natural world. It seems clear that the use of the term "value" has undergone many developments and has come to be used to mediate different functions and to convey other information, and yet, pleasure seems to be the *source,* as it where of these functions and uses. I considered to withdraw to the position that pleasure is best understood as *proto-value*, some sort of developmental pre-cursor to the full-fledged concept as used in everyday talk, but eventually decided against it. It seems to be part of what make philosophy interesting that we look for what is most fundamental about the notion under investigation. We want to know what the *essence* of reality, time, and consciousness is, and value-theory is no different. Of course, sometime this search lead to the realisation that there is no "essence", that the notion under investigation cover quite distinct phenomena, or a phenomena with blurred boundaries. But if there is a common origin of value discourse, a discourse that might then have become quite complex, this is where we should go look for the nature of value. More than an argument, this is a naturalistic *credo*.

I've opted out of some big debates in meta-ethics, by treating this discipline less as a debate and more as a number of theoretical *decisions*. There is an end to the type of argument that every participant in the debate have reason to accept. Even if every now and then someone points out a logical flaw our implausibility in the other side, this rarely provides decisive reasons to abandon any fundamental approach, only to make it more specific. It seems to me that naturalists have reasons to accept a different methodology from non-naturalists, and I've no idea how to provide a knock-down argument on either side. The

account is furnished with conditionals about theoretical claims, and for most of them, I believe there are considerations in favour of the route I've chosen, but it would take a much more ambitious meta-theory to undertake to argue conclusively for all those claims.

Things that we believe to be true about value, the facts that make us suspect that there is such a thing at all, are reducible to, or explainable in terms of, pleasure and hedonic processes. Does this mean that hedonism is true? Ah, now that's the pickle. A theory is true only if the claims it is making are true, and I'm construing hedonism as making only those claims.

The truth of hedonism is a truth about value because the place of pleasure in the explanation of evaluation shows it to play a fundamental role, more so than any other candidate property present in these explanations. This hedonism is compatible with our evaluations, desires and preferences picking out a pluralistic set of objects. In one sense, I'm not committed to such evaluations being *mistakes*. The argument for hedonism is a psychological one, but it is not the argument that we would value pleasure and only pleasure if we were well informed and freed of cognitive infirmities: it is not a theory about what we *would* value, but about the *causal explanation* of our evaluations.

Emotions are evaluating because they are valenced, and they are valenced because they incorporate a value on the hedonic-doloric scale. Revealing that evaluative beliefs depend on affective processes might undermine our confidence in these evaluations, and there are circumstances were it should. If we find that there is no tendency for these emotional reactions to reliably detect any objective feature, or for informed and rational people to converge in their affective reactions, it's hard to assign them status as evidence for anything. Instead, it would favour the interpretation that evaluations involve some sort of mistakes. But since what is *being* manipulated in these scenarios is the hedonic system, this doesn't undermine our beliefs in the value of *pleasure*. Uniquely, then, our hedonic values survive this test.

The theory of value is fraught with controversies, with fundamental disagreements about what it should do, what is required from the property if, indeed, it is a property at all. All this makes it difficult and, in fact, ill advised to make decisive claims about the truth of ones theory. Hedonism, I've argued, should be engaged in an explanatory project, one that engages with empirical science as well as with conceptual analysis. That is how it can substantiate it claims. But does it mean that pleasure is the good? I prefer to put it this way:

Pleasure is the good *enough.*

Bibliography:

Adams, Robert Merrihew, (1999): *Finite and Infinite Goods: A Framework for Ethics*, Oxford University Press, New York.

Alston, William P. (1967): *Pleasure*, in Paul Edwards (ed.) The Encyclopedia of Philosophy, Macmillian, New York 341-7

Anscombe, G.E.M. (1958): *Modern Moral Philosophy*, Philosophy vol 33, 1-19

Anscombe, G.E.M. (1967): *On the grammar of enjoy*, Journal of Philosophy vol. 64, 607-614

Appiah, Anthony Kwame: *Experimental ethics*, Harvard University Press, Cambridge, MA

Aristotle (1999): *Nicomachean Ethics*, 2nd edition, Hacket Publishing Company, Indianapolis

Aydede, Murat (2000): A theory of Pleasure vis-á-vis Pain, Philosophy and Phenomenological Research 61 537-70

Aydede, Murat (2005): *The Main Difficulty with Pain*, in Murat Aydede ed. (2005), 123-136

Aydede, Murat. (ed) (2005): *Pain: New Essays on Its Nature and the Methodology of Its Study*, A Bradford Book, the MIT Press, Cambridge, Mass

Ayer, A.J. (2001): *Language, Truth and Logic*, Penguin Books, London

Ball, Stephen W.(1991): *Linguistic Intuitions and Varieties of Ethical Naturalism*, Philosophy and Phenomenological Research, Vol 51, No 1 1-38

Barnett, David (2002): *Against* A Posteriori *Moral Naturalism*, Philosophical Studies, vol. 107, 239-257

Bengtsson, David (2003a): *The Intrinsic Value of Pleasure*, in Rabinowicz and Rønnow-Rasmussen (eds.) *Patterns of Value: Essays on Formal Axiology and Value Analysis*, Lund Philosophy Reports 2003:1, 29-61

Bengtsson, David (2003b): *The Nature of Explanation in a theory of consciousness*, Lund University Cognitive Studies 106

Bengtsson, David (2004): *Pleasure and the Phenomenology of Value*, in Rabinowicz and Rønnow-Rasmussen (eds.) *Patterns of Value: Essays on Formal Axiology and Value Analysis, II*, Lund Philosophy Reports 2004:1, 21-35

Bentham, Jeremy (1789): *An Inquiry into the principle of morals and legislation*,, London

Bentham, Jeremy (1960): *Introduction to the principles of morals and legislation* ed. W Harrison, Blackwell, Oxford

Berridge, Kent C. (1999): *Pleasure, Pain, Desire and Dread: Hidden Core Processes of Emotion*,, in Kahneman, Diener and Schwarz (1999) 525-557

Berridge, Kent C. (2003): *Pleasures of the brain*, Brain and Cognition, Vol. 52, 106-128

Berridge, Kent C. (2004): *Motivation concepts in behavioural neuroscience*, Physiology & Behavior, Vol. 81, 179-209,

Blackburn, Simon (1993): *Essays in Quasi-Realism*, Oxford University Press, New York

Block, Ned (1995), *On a Confusion about a Function of Consciousness*, Behavioral and Brain Sciences 18, 227-47,

Block, Ned (2005): *Bodily sensation as an obstacle for Representationism*, in Aydede (2005) 137-142

Boyd, Richard (1988): *How to be a moral realist*, in Sayre-McCord (1988), 188-227

Boyd, Richard (2003): *Finite Beings, Finite Goods: The Semantics, Metaphysics and Ethics of Naturalist Consequentialism, Part 1*, Philosophy and Phenomenological Research, vol. 66. No. 3, 505- 554

Bradley, F.H. (1962), *Ethical Studies*, 2nd edition, Oxford University Press, Oxford

Brandt, Richard B. (1954): *Hopi Ethics*, University of Chicago Press, Chicago

Brandt, Richard B. (1959): *Ethical Theory: The Problems of Normative and Critical Ethics*,: Prentice-Hall, Englewood Cliffs, N.J.

Brandt, Richard B. (1967): *"Hedonism,"* in Paul Edwards (ed.) The Encyclopedia of Philosophy, vol 4., Macmillan, Free Press, New York, 432-5

Brandt, Richard (1985), *The Explanation of Moral Language,* in Copp and Zimmerman (1985)

Brandt, Richard B. (1998): *A Theory of the Good and the Right*, Prometheus Books, Amherst, NY

Brax, David (2009), *Estetiskt värde, värderande responser* , 326-42 in Mortensen, Anders (ed.) *Litteraturens värden,* Brutus Östlings Förlag Symposion

Bressan, R. A. and Crippa, J. A. (2005): *The role of dopamine in reward and pleasure behavior – review of data from preclinical research*, Acta Psychiatrica Scandinavica, Vol. 111 (Suppl. 427), 14-27

Brink, David (1989), *Moral Realism and the Foundation of Ethics*, Cambridge University Press, Cambridge

Cambell, Richmond and Woodrow, Jennifer (2003): *Why Moore's Open Question is Open: the Evolution of Moral Supervenience*, The Journal of Value Inquiry 37, 353-372

Chalmers, David (1996): *The Conscious Mind*, Oxford University Press, New York

Chalmers, David and Jackson, Frank (2001): *Conceptual Analysis and Reductive Explanation*, Philosophical Review 110, 315-361

Chalmers, David (2004): *The Representational Character of Experience*, in ed. Leither, Brian, *The Future of Philosophy*, Oxford University Press, New York

Chalmers, David (2006) *Two-dimensional Semantics* in Lepore and Smith (eds.) *Handbook of the Philosophy of Language*, Oxford University Press, New York

Chisholm, Roderick M (1986). *Brentano on Intrinsic Value*, Cambridge University Press, Cambridge

Clore, Gerald L. and Huntsinger, Jeffrey R.: *How emotions inform judgment and regulate thought*, Trends in Cognitive Sciences, Vol. 11, No. 9, 393-399 (2007)

Copp, David and Zimmerman, David (eds.) (1985): *Morality, Reason and Truth*, Rowman And Allanheld, Totowa, N.J

Copp, David (1990): *Explanation and Justification in Ethics*, Ethics 100, 237-258

Copp, David (2000): *Milk, Honey, and the Good Life on Moral Twin Earth*, Synthese 124, 113-137

Copp, David (2003): *Why Naturalism?,* Ethical Theory and Moral Practice 6: 179–200

Crisp, Roger (2005): *Value, reasons and the structure of justification: how to avoid passing the buck,* Analysis, Vol 65, No. 1 80-85

Crisp, Roger (2006): *Reasons & the Good*, Clarendon Press, Oxford

Cullity, Garrett (2006): *As You Were? Moral philosophy and the aetiology of moral experience*, Philosophical Explorations, Vol. 9. No 1. 117-131

Damasio, Antonio (1994): *Descartes Error*, Quill, Harper Collins

Dancy (1982) 'Ethical Particularism and Morally Relevant Properties,' Mind, vol. 92, 530-47

Daniels, Norman (1979), "Wide Reflective Equilibrium and Theory Acceptance in Ethics," Journal of Philosophy, 76: 256–282

D'Arms, Justin and Jacobson, Daniel (2000): *Sentiment and Value*, Ethics 110, 722-748

Darwall, Stephen; Gibbard, Allan and Railton, Peter: (1992) *Toward* Fin de Siecle *Ethics: Some Trends*, The Philosophical Review, Vol. 101, No 1. 115-189

Darwall, Stephen (2006): *How Should Ethics Relate to (the Rest of) Philosophy? Moore's Legacy*, in ed. Horgan, Terry and Timmons, Mark: *Metaethics after Moore*, Oxford University Press, New York, 17-37

Davidson, Richard (1994): *On Emotion, Mood and Related Affective Constructs,* in Ekman and Davidson (eds.) (1994)

Davidson, Richard J.: Jackson, Daren C.: and Kalin, Ned H. (2000): *Emotion, Plasticity, Context, and Regulation: Perspectives From Affective Neuroscience*, Psychological Bulletin, Vol. 126, No. 6, 890-909

Davidson, Richard J.; Scherer, Klaus R. and Goldsmith, H. Hill (eds.) (2003) *Handbook of Affective Sciences*, Oxford University Press, New York

Dennett, Daniel (1978): *Why you can't make a computer that feels pain*, Synthese Volume 38, number 3, 415-56

De Waal, Frans (1996): *Good-natured – the Origins of Right and Wrong in Humans and Other Primates*, Harvard University Press, Cambridge, MA

Donellan, Keith S. (1966): *Reference and Definite Descriptions*, The Philosophical Review, Vol. 75, No. 3. 281-304

Doris, John M. and Stich, Stephen P. (2006): *As a Matter of Fact: Empirical Perspectives on Ethics,* Philosophical Explorations, Vol. 9. No 1., 114-152

Douglas, Guy (1998): *Why pains are not mental objects,* Philosophical Studies, Vol. 91, No 2, 127-48

Duncker, Karl (1941) *On Pleasure, Emotion and Striving*, Philosophy and Phenomenological Research vol. 1, 391-430

Ekman, Paul and Davidson, Richard J. (eds.) (1994): *The Nature of Emotion Fundamental Questions*, Oxford University Press, New York

Ellsworth, Phoebe C., and Scherer, Klaus R.(2003): *Appraisal Processes in Emotion*, in Davidson, Scherer and Goldsmith, 572-595

Epstein, Seymor (1994): *Integration of the Cognitive and the Psychodynamic Unconscious*, American Psychologist vol. 49, No. 8, 709-724

Ewing, A.C. (1939): *A Suggested Non-Naturalistic analysis of the good* , Mind New Series, Vol. 48, No. 189, 1-22

Feldman, Fred (1997a): *the Intrinsic Value of Pleasure* Ethics, Vol. 107, No. 3, 448-466

Feldman, Fed (1997b): *Two questions about pleasure*, in *Utilitarianism, Hedonism and Desert: Essays in Moral Philosophy*, Cambridge University Press, Cambridge

Feldman, Fred (2004): *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*, New York: Oxford University Press

Field, Harry (1973): *Theory Change and the Indeterminacy of Reference*, The Journal of Philosophy, Vol. 70, No. 14, 462-481

Flanagan, Owen (1998): *Ethics Naturalized: Ethics as Human Ecology* in May, Friedman and Clark, 19-44

Foot, Philippa: *Does Moral Subjectivism Rest on a Mistake?*, Oxford Journal of Legal Studies, Vol. 15, No 1. 1-14 (1995)

Foot, Philippa: *Natural Goodness*, Oxford University Press (2001)

Forgas, Joseph P. (2003): *Affective Influences on Attitudes and Judgments*, in Davidson, Scherer and Goldsmith, 596-618

Forster, E.M.(2006): *The Longest Journey,* Penguin Classics, London

Frankena, W.K (1939).: *The Naturalistic Fallacy*, Mind, vol. 48, No 192, 464-477

Frankena, W.K. (1958): *Obligation and motivation*, in *Essays in Moral Philosophy,* Melden (ed.), University of Washington Press, Seattle, A.I., 40-81

Frankena, W.K. (1973); *Ethics*, Prentice-Hall, Englewood Cliffs, N.J.:

Gibbard, Allan (2003): *Thinking How to Live*, Harvard University Press, Cambridge, MA

Gibbard, Allan (2006): *Normative Properties*, in Horgan, Terry and Timmons, Mark (eds.): *Metaethics after Moore*, Oxford University Press, New York, 319-338

Goldman, Alan H (1987): *Red and Right*, The Journal of Philosophy, vol. 84, No. 7, 349-362

Goldman, Alan H (1990): *Aestethic qualities and aesthetic value*, Journal of Philosophy vol. 87, No. 1, 23-37

Goldman, Alvin I. (2007): Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status, Grazer Philosophische Studien vol. 74, 1–26

Goldstein, Irwin (1989): *Pleasure and Pain: Unconditional, Intrinsic Values*, Philosophy and Phenomenological Research, volume 50, 255-276

Gosling, J.C.B.: (1969) *Pleasure and Desire – the Case for Hedonism Reviewed,* Clarendon Press, Oxford

Greene, Joshua, et al (2001): *An fMRI Investigation of Emotional Engagement in Moral Judgement*, Science, Vol. 293, 2105-2108

Greene, Joshua and Haidt, Jonathan (2002): *How (and where) does moral judgment work?* in Trends in Cognitive Sciences vol. 6, 517-523

Greene, Joshua, et al (2004): *The neural bases of cognitive conflict and control in moral judgment* in Neuron vol. 44, 389-400

Griffin, James (1986): *Well-being: It's Meaning, Measurement, and Moral Importance,* Clarendon Press, Oxford

Haidt, Jonathan (2001): *The emotional dog and its rational tail*, Psychological review Vol. 108, 814-834

Haidt, Jonathan (2006): *The Happiness Hypothesis*, Arrow Books, London

Hall, Richard J. (1989) *Are pains necessarily unpleasant,* Philosophy and Phenomenological Research, Vol. 49, No. 4, 643-659

Hare, R.M. (1952): *The Language of Morals*, Clarendon Press, Oxford

Hare, R.M. (1957): *Are Discoveries about the Uses of Words Empirical?* The Journal of Philosophy, Vol. 54, No. 23, 741-750

Hare, R.M. (1981): *Moral Thinking*, Oxford University Press, Oxford

Hare, R.M.(1993): *Objective Prescriptions*, Philosophical Issues, Vol 4, Naturalism and Normativity, 15-32

Harman, Gilbert (2000): *Explaining Value and Other Essays in Moral Philosophy*, Clarendon Press, Oxford

Harman, Gilbert, (1977) *The Nature of Morality*, Oxford University
Press, New York

Harman, Gilbert (1986): *Moral Explanations of Natural Facts: Can Moral Claims Be Tested Against Reality?*, Southern Journal of Philosophy, 24 (Supplement): 57–68

Hauser, Mark (2007): *Moral Minds,* Harper Perennial, New York

Heathwood, Chris (2006): Desire-satisfactionism and Hedonism, Philosophical Studies 128, 539-563

Heathwood, Chris (2007): The reduction of sensory pleasure to desire, Philosophical Studies 133, 23-44

Helm, Bennett W. (2001): *Emotional Reasons – Deliberation, Motivation, and the Nature of Value*, Cambridge Studies in Philosophy, Cambridge

Helm, Bennett W.:(2002) *Felt Evaluations: A Theory of Pleasure and Pain*, American Philosophical Quarterly, 39:13-30

Horgan, Terence and Timmons, Mark (1991): *New Wave Moral Realism Meets Moral Twin Earth*, Journal of Philosophical Research Vol. 16, 447-466

Horgan, Terry and Timmons, Mark (eds) (2006): *Metaethics after Moore*, Oxford University Press, New York

De Houwer, Jan; Thomas, Sarah and Baeyens, Frank (2001): *Associative Learning of Likes and Dislikes: A Review of 25 Research on Human Evaluative Conditioning*, Psychological Bulletin, Vol. 127 No 6 853-869

De Houwer, Jan (2007): *A Conceptual and Theoretical Analysis of Evalautive Conditioning*, The Spanish Journal of Psychology, Vol 10, No. 2, 230-241

Hume, David (1978): *A treatise of Human Nature*, ed. Selby-Bigge & Niddich, 2nd edition, Oxford University Press, Clarendon

Hume, David (1874-75): *On the standard of taste,* in T.H Green och T.H Grose (red.), Longman and Green, London

Jackson, Frank and Pettit, Philip (1995), *Moral Functionalism and Moral Motivation*, The Philosophical Quarterly vol. 45, No 178 20-40

Jackson, Frank and Pettit, Philip (1996): *Moral Functionalism, Supervenience and Reductionism*, The Philosophical Quarterly, Vol. 46, No. 182, 82-86,

Jackson, Frank (1998): *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford University Press, New York

Jackson, Frank, Pettit, Philip and Smith, Michael (2004): *Ethical Particularism and Patterns*, in *Mind, Morality and Explanation*, Oxford University Press, New York

Jackson, Frank (1982) *Epiphenomenal Qualia*' Philosophical Quarterly no 32, 127-136

Jackson, Frank (1974): *Defining the autonomy of ethics* *The Philosophical Review*, Vol. 83, No. 1, 88-96

Jackson, Frank (2003): *Cognitivism, A Priori Deduction, and Moore*, Ethics, 113: 557-75

James, William (1950): *The Principles of Psychology*, Dover Publications, Inc., New York

Johnston, Mark (2001): *The Authority of Affect,* Philosophy and Phenomenological Research, Vol. 63, No. 1, 181-214

Jones, Christopher R., Fazio Russell H., and Olson, Mikael A. (2009): *Implicit Misattribution as a Mechanism Underlying Evaluative Conditioning*, Journal of Personality and Social Psychology Vol. 96, No. 5, 933–948

Joyce, Richard (2006): *Meta-ethics and the empirical sciences*, Philosophical Explorations vol. 9 133-147

Joyce, Richard (2007): *the Evolution of Morality*, A Bradford Book, the MIT Press, Cambridge, MA

Joyce, Richard (2008): *What Neuroscience can (and Cannot) Contribute to Metaethics*, in Sinnott-Armstrong (ed.) (2008)

Kagan, Shelly (1998): *Rethinking Intrinsic Value*, The Journal of Ethics vol. 2, 277-297

Kagan, Shelly (1992): *The Limits of Well-being*, Social Philosophy and Policy 9, 169-89

Kahneman Daniel; Wakker, Peter P.; Sarin, Rakesh (1997): *Back to Bentham? Explorations of Experienced Utility*, The Quarterly Journal of Economics, May 375-405

Kahneman, Daniel (1999): *Objective Happiness*, in Kahneman, Diener and Schwarz (1999) 3-25

Kahneman, Daniel; Diener, Ed and Schwarz, Norbert (eds) (1999): *Well-Being – The Foundations of Hedonic Psychology*, Russell Sage Foundation, New York

Kahneman, Daniel and Tversky, Amos (eds.): *Choices values and frames*, Cambrdige University Press, Cambridge

Katz, Leonard D.,(1986): *Hedonism as Metaphysics of Mind and Value*, UMI Disertation Services

Katz, Leonard D. (2006), "Pleasure", *The Stanford Encyclopedia of Philosophy (Summer 2006 Edition)*, Edward N. Zalta (ed.), forthcoming

URL = <http://plato.stanford.edu/archives/sum2006/entries/pleasure/>.

Katz, Leonard D. (2008): *Hedonic Reasons as Ultimately Justifying and the Relevance of Neuroscience*, in Sinnott-Armstrong (ed.)

Kawall, Jason (1999): *The Experience Machine and Mental State Theories of Well-being*, The Journal of Value Inquiry, vol. 33, 381-38

Kim, Jaegwon (1997): *Moral Kinds and Natural Kinds: What's the Difference – for a Naturalist?*, Philosophical Issues, Vol. 8, 293-301

Korsgaard, Christine (1983): "Two Distinctions in Goodness", Philosophical Review, vol. 92, 169-95

Kringelbach, Morten L. *The Human Orbitofrontal Cortex: Linking Reward to Hedonic Experience*, Nature Reviews Neuroscience, Vol. 6, 691-701, (2005)

Kringelbach, Morten L. (2009) *The pleasure center*, Oxford University Press, New York

Kripke, Saul (1972): *Naming and Necessity*, in Davidson and Harman (eds.) *Semantics of Natural Language, Synthese* Library, Volume 40

Kölbel, Max (2004): *Faultless Disagreement*, Proceeding of the Aristotelian Society, Vol. 104, Issue 1, 53- 73 (2004)

Layard, Richard (2005): *Happiness – Lessons From a New Science*, The Penguin Press, New York (2005)

LeDoux, Joseph (1996): *The Emotional Brain - The Mysterious Underpinnings of Emotional Life*, Simon and Schuster Paperbacks, New York

Lemos, Noah (1994): *Intrinsic Value*, Cambridge University Press, Cambridge

Lenman, James, "Moral Naturalism", *The Stanford Encyclopedia of Philosophy (Fall 2006 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2006/entries/naturalism-moral/>.

Lepper, Mark; Greene, David and Nisbett, Richard E. (1973): Undermining Children's Intrinsic Interest with Extrinsic Reward: A test of the "Overjustification" Hypothesis, Journal ol Personality and Social Psychology, Vol. 28, No. 1, 129-137

Lewis, C. I. (1946): *An Analysis of Knowledge & Valuation*, Open Court, La Salle

Lewis, David (1970): *How to Define Theoretical Terms*, Journal of Philosophy, 67, 427-446

Lewis, David (1972): *Psychophysical and theoretical identifications*, the Australasian Journal of Philosophy 50, 249-258

Lewis, David (1986): *On the plurality of Worlds*, Blackwell Publishers, Oxford

Lewis, David (1989): *Dispositional theories of value*, Proceeding of the Aristotelian Society, Suppl. Vol. 63, 113-137

Locke, John (1975): *An Essay Concerning Human Understanding*, ed. P Nidditch, Clarendon Press, Oxford

Loewenstein, George and Lerner, Jennifer S.: *The Role of Affect in Decision Making*, in Davidson, Scherer and Goldsmith (2003) 619-642

Mackie, J.L.(1977): *Ethics: Inventing Right and Wrong*,, Penguin, Harmondsworth

Marshall, Henry Rutgers (1892): *the Field of Aesthetics Psychologically Considered*, Mind New Series, Vol 1. No. 3, 358-378

May, Larry; Friedman, Marilyn and Clark, Andy (eds.) (1998): *Mind and Morals – Essays on Ethics and Cognitive Science*, A Bradford Book, the MIT Press, Cambridge, MA

Maund, Barry (2005): *Michael Tye on Pain and Representational Content*, in Aydede (ed)

McDowell, John (1988): *Values and Secondary Qualities*, in Sayre-McCord (1988), 166-180

McGinn, Colin (1976): *On the Necessity of Origin*, The Journal of Philosophy, Vol. 73. No. 5, 127-135

McNaugton, David and Rawling, Piers (2003): *Naturalism and Normativity,* Aristotelian Society Supplementary Volume, vol. 77, issue 1, 23-45

Melzack, R. and Casey, K.L. (1968): *Sensory, Motivational, and Central Control Determinants of Pain: A New Conceptual Model*, in D Kenshalo, (Ed) *The Skin Senses*: Charles C. Thomas, Springfield 223–43.

Melzack and Wall (1965): *Pain Mechanisms: a new theory*, Science vol. 150, Nov 19, 971-9

Mendola, Joseph (1990): *Objective Value and Subjective States*, Philosophy and Phenomenological Research, Vol. 50, Issue 4, 695-714

Mendola, Joseph (2006): *Intuitive Hedonism*, Philosophical Studies Vol. 128, 441-477

Mill, John Stuart (1993): *Utilitarianism, on Liberty, Considerations on Representative Government*, The Everyman Library, New York

Mill, John Stuart (1950): *of propositions merely verbal*, in ed. Ernest Nagel *John Stuart Mill's Philosophy of Scientific Method*, Hafner Press, New York

Millgram, Elijah (2000): *Mill's Proof of the Principle of Utility,* Ethics, vol. 110, No. 2, 282-310

Momeyer, Richard W (1975): *Is Pleasure a Sensation?*, Philosophy and Phenomenological Research vol. 36:113-21

Moore, G.E. (1993), *Principia Ethica*, Revised Edition ed. Thomas Baldwin, Cambridge University Press, Cambridge

Much, Jochen and Klauer, Karl Christoph (eds.) (2003): *The psychology of evaluation*, Erlbaum, Mahway, NJ

Nagel, Thomas (1986): *The View from Nowhere*, Oxford University Press, New York

Nagel, Thomas (1974), *What is it like to be a bat?* The Philosophical Review Vol. 84, issue 4 435-50

Nettle, Daniel (2005): *Happiness – The Science Behind Your Smile*, Oxford University Press, New York

Nichols, Shaun (2004): *Sentimental Rules – On the Natural Foundations of Moral Judgement*, Oxford University Press, New York

Nozick, Robert (1974): *Anarchy State and Utopia,* Basil Blackwell Press, Oxford

Oddie, Graham (2005): *Value, Reality, and Desire*, Oxford University Press, New York

Panksepp, Jaak (1998): *Affective Neuroscience – the Foundations of Human and Animal Emotions*, Oxford University Press, New York

Papineau, David (1993): *Philosophical Naturalism*, Blackwell, Oxford

Parfit, Derek (1984): *Reasons and Persons*, Oxford University Press, New York

Petts, Jeffrey (2000): *Aesthetic Experience and the Revelation of Value*, The Journal of Aesthetics and Art Criticism, Vol. 58, No. 1, 61-71

Perry, David L.,(1967): *The Concept of Pleasure*, Mouton, The Hague

Persson, Ingmar (2005): The Retreat From Reason, Oxford University Press, Oxford

Plato (1975): *Philebus*, Gosling (red.) Oxford University Press, London

Prinz, Jesse (2006): *The Emotional Basis of Moral Judgments*, Philosophical Explorations, Vol. 9, No 1, 29-43

Putnam, Hilary (1973): *Meaning and Reference*, The Journal of Philosophy,Vol. 70, No. 19, 699-711

Putnam, Hilary (1981): *Reason, Truth and History*, Cambridge University Press, Cambridge

Quine, W v O (1969): *Epistemology naturalized*, in *Ontological Relativity and Other Essays*,, Columbia University Press, New York

Quine, W v O and Ullian, J.S. (1978) : *The Web of Belief*, McGraw-Hill, New York

Rabinowicz, Wlodek and Rønnow-Rasmussen, Toni (1999): *A distinction in Value: Intrinsic and For Its Own Sake*, Proceeding of the Aristotelian Society, 100, 33-52

Rabinowicz, Wlodek and Rønnow-Rasmussen, Toni (2004): *The Strike of the Demon*: *On Fitting Pro-attitudes and Value*, Ethics, 114, 391-423

Rabinowicz, Wlodek and Österberg, Jan (1996): *Value Based on Preferences: ON Two Interpretations of Preference Utilitarianism*, Economics and Philosophy, Vol 12, No. 1, 1-27

Rachels, Stuart (2004): *Six theses about pleasure*, Philosophical Perspectives, Vol 18, 247-67.

Rachels, Stuart (2000) *Is Unpleasantness Intrinsic to Unpleasant Experiences?* Philosophical Studies, Vol. 99, No. 2., 167-210

Railton, Peter (1989): *Naturalism and prescriptivity*, Social Philosophy and Policy, vol. 95, 151-174

Railton, Peter (1998): *Moral Explanation and Moral Objectivity*, Philosophy and Phenomenological Research vol. 58, 175-182

Railton, Peter (2003: *Facts, Values and Norms*, Cambridge University Press, Cambridge

Rawls, John (1971): *A theory of Justice,* Harvard University Press, Cambridge MA

Van Roojen, Mark (1996): *Moral Functionalism and Moral Reductionism*, The philosophical Quarterly, Vol. 46 No. 182 77-81

Rosati, Connie S. (1995): *Naturalism, Normativity, and the Open Question Argument*, Nous, Vol 29, No. 1, 46-70,

Ross, David (2002): *The Right and the Good*, Oxford University Press, New York

Rubin, Michael (2008): *Sound Intuitions on Moral Twin Earth* , Philosophical Studies, Vol. 139, No. 3, 307-27

Russell, James A. (2003): *Core Affect and the Psychological Construction of Emotion,* Psychological Review vol. 110, No. 1, 145-72.

Ryan, Richard M. and Deci, Edward L. (2001): *On Happiness and Human Potentials: A review of research on hedonic and eudaimonic Well-being*, Annual Review of Psychology, Vol. 52, 141-166

Ryle, Gilbert (1969): *Dilemmas*, Cambridge University Press, Cambridge

Ryle, Gilbert, (2000) *The Concept of Mind*, Penguin Classics, London

Rønnow-Rasmussen, Toni (2002). *Hedonism, Preferentialism, and Value Bearers,* the Journal of Value Inquiry*,* vol.36, 463-472

Sayre-McCord, Geoffrey, (ed.) (1988), *Essays on Moral Realism*, Cornell University Press, Ithaca

Sayre-McCord, Geoffrey (1988): *Moral Theory and Explanatory Impotence,* in Sayre-McCord (1988) 256-281

Sayre-McCord, Geoffrey (1991): *Being a Realist About Relativism (in Ethics)*, Philosophical Studies 61, 155-176

Sayre-McCord, Geoffrey (1997): *'Good' on Twin Earth*, Philosophical Issues, vol. 8, 267-292

Sayre-McCord, Geoffrey (2001): *Mill's "Proof" of the Principle of Utility: A More than Half-Hearted Defense*, Social Philosophy & Policy, vol. 18, no 2, 330-360.

Scanlon, T.M. (2000): *What We Owe to Each Other*, Belknap, Harvard University Press, Cambridge, MA

Schroeder, Timothy (2004):*Three Faces of Desire*, Oxford University Press, New York

Schwartz, Norbert and Clore, Gerald L. (1983) *Mood, Misattribution, and Judgments of Weil-Being: Informative and Directive Functions of Affective States*, Journal of Personality and Social Psychology, Vol. 45, No. 3, 513-523

Schwartz, Norbert and Clore, Gerald, L. (2003): *Mood as Information: 20 Years Later,* Psychological Inquiry, Vol. 14, No. 3&4, 296–303

Schultz, Wolfram: *Multiple Reward Signals In the Brain* (2000), Nature Reviews Neuroscience, Vol. 1, 199-207

Shafer-Landau, Russ (2006): *Ethics as Philosophy: A Defence of Ethical Nonnaturalism*, in Horgan, Terry and Timmons, Mark (eds.): *Metaethics after Moore*, Oxford University Press, New York, 209-232

Sidgwick, Henry (1892): *The Feeling-Tone of Desire and Aversion,* Mind, New Series, Vol 1. No 1. 94-101

Sidgwick, Henry (1981): *The Methods of Ethics*, 7th edition, Hackett Publishing Company, Indianapolis, IN

Silverstein, Matthew (2000): *In defense of happiness: a response to the experience machine*, vol. 26 nr. 2, 279-301

Sinnott-Armstrong, Walter (ed.) (2008): *Moral Psychology, vol 3.* the MIT Press, Cambridge, MA.

Slovic, Paul; Finucane, Melissa L.; Peter, Ellen; MacGregor, Donald G (2007): *The affect heuristic*, European Journal of Operational Research, vol. 177, 1333–1352

Slote, Michael (1992): *Ethics Naturalized*, Philosophical Perspectives, Vol. 6, Ethics, 355-65

Smith, Michael (1994): *The Moral Problem*, Blackwell Publishing, Oxford

Smith, Michael (2003): *Neutral and Relative Value after Moore*, Ethics, Vol. 113, 576-598

Smith, Michael (2004): *Ethics and the a Priori – selected essays on moral psychology and meta-ethics*, Cambridge University Press, Cambridge

Sobel, David (1999): *Pleasure as a mental state*, Utilitas, vol. 11, 230-4

Sobel, David (2002): *Varieties of Hedonism*, Journal of Social Philosophy, Vol. 3. No. 2, 240-256

Solomon, Robert M. (2003): *Emotions, thoughts and feelings: What is a 'cognitive theory' of emotions and does it neglect affectivity?*, in Hatzimousis, Anthony (ed.) *Philosophy and the Emotions,* Cambridge University Press, Cambridge, 1-18

Sosa, Ernest (1997): *Water, Drink, and "Moral Kinds"*, Philosophical Issues, Vol. 8, 303-312

Sprigge, T.L.S. (2000): *Is the Esse of Intrinsic Value Percipi? Pleasure, Pain and Value*, In Anthony O'Hear (ed) *Philosophy, the Good, the True and the Beautiful*, Cambridge University Press, Cambridge, 119-140

Sterne, Lawrence (2004): *The Life and Opinions of Tristram Shandy, Gentleman,* Modern Library, New York

Stevenson, C.L. (1937): *The emotive theory of ethical terms*, *Mind*, New Series, Vol. 46, No. 181, 14-31

Street, Sharon (2006): *A Darwinian Dilemma for Realist Theories of Value*, Philosophical Studies 127; 109-166

Sturgeon, Nicholas L.(1985): *Moral Explanations*, in Copp and Zimmerman, 49-78

Sturgeon, Nicholas L.(2002): *Ethical Intuitionism and Ethical Naturalism*, in Stratton-Lake (2002) 184-211

Sturgeon, Nicholas L.(2003): *Moore on Ethical Naturalism*, Ethics 113, 528-556

Sturgeon, Nicholas L.(2005): *Ethical Naturalism*, in David Copp (ed), *The Oxford Handbook of Ethical Theory*, Oxford University Press, New York, 91-120

Stratton-Lake, Philip (2002): (ed.) *Ethical Intuitionism: Re-evaluations*, Clarendon Press, Oxford

Strawson, Galen (1994), *Mental Reality*, The MIT Press, Cambridge, MA

Sumner, Wayne (1996): *Well-fare, Happiness and Ethics,* Clarendon Press, Oxford

Tersman, Folke (1993): *Reflective Equilibrium* – an Essay in Moral Epistemology, Stockholm Studies in Philosophy 14, Almqvist&Wiksell International

Tye, Michael (2005): *Another look at representationalism about pain,* in Aydede (2005) 99-120

Weisberg et al. (2008): *The seductive allure of neuroscience explanations*, Journal of Cognitive Neuroscience, vol. 20, 470-477

Wiggins, David (1987): *A Sensible Subjectivism*, in *Needs, Values, Truth*, Blackwell, Oxford, 185-214

Williams, Bernhard (2006): *Ethics and the Limits of Philosophy*, Routledge, Oxford

Williams, Laurence and Barg, John (2008): *Experiencing physical warmth promotes interpersonal warmth*, Science vol .322, 606-609

Zajonc, Robert (1980): Feeling and thinking: Preferences need no inferences, *American Psychologist*. Vol. 35, No. 2, 151-175

Zangwill, Nick (2000): *Against Analytic Moral Functionalism*, Ratio 13, 275-286

Zimmerman, Michael J (1980). 'On the Intrinsic Value of States of Pleasure', *Philosophy and Phenomenological Research*, Volume 41, 26-45

Zimmerman, Michael (2001), *the Nature of Intrinsic Value,* Rowman and Littlefield Publishers, Lanhamn, MD